

This Page Is Inserted by IFW Operations
and is not a part of the Official Record

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

**As rescanning documents *will not* correct images,
please do not report the images to the
Image Problem Mailbox.**

PATENT ABSTRACTS OF JAPAN

(11) Publication number: **09050527 A**

(43) Date of publication of application: **18.02.97**

(51) Int. Cl. **G06T 7/00**
G06K 9/20

(21) Application number: 07203259

(22) Date of filing: 09.08.95

(71) Applicant: **FUJITSU LTD**

(72) Inventor: OBARA ATSUKO
NAOI SATOSHI

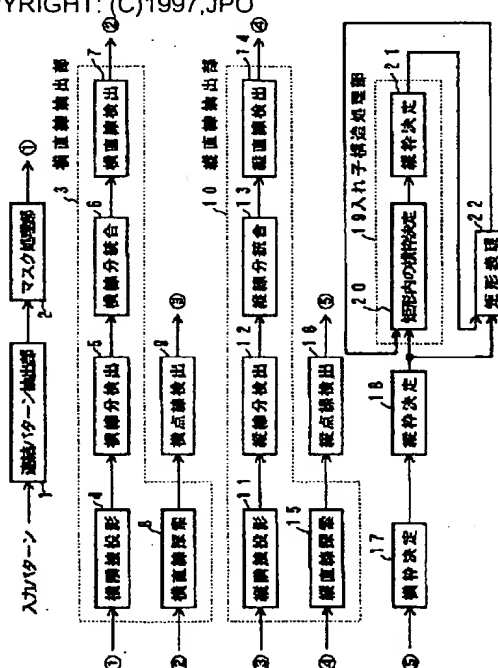
(54) FRAME EXTRACTING DEVICE AND RECTANGLE
EXTRACTING DEVICE

COPYRIGHT: (C)1997,JPO

(57) Abstract:

PROBLEM TO BE SOLVED: To accurately extract a frame out of table format frames having rectangular structure in which the external shapes of frames are irregular even when a character is in contact with a frame or a character is extruded from a frame, to accurately extract the frame.

SOLUTION: A horizontal straight line extracting part 3 and a vertical straight line extracting part 10 extract respectively straight lines from a partial pattern extracted by a connected pattern extracting part 1, a horizontal frame determining part 17 determines horizontal frames by combining each two horizontal straight lines extracted by the extracting part 3 successively from the uppermost one, a vertical frame determining part 18 determines vertical frames from each two vertical straight lines extracted by the extracting part 10, and a rectangle expressing part 22 extracts a rectangular range surrounded by two horizontal frames and two vertical frames as a frame.



(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号

特開平9-50527

(43) 公開日 平成9年(1997)2月18日

(51) Int.Cl. ⁸	識別記号	庁内整理番号	F I	技術表示箇所
G 0 6 T 7/00		9061-5H	G 0 6 F 15/70	3 3 0 L
G 0 6 K 9/20	3 4 0		G 0 6 K 9/20	3 4 0 J

審査請求 未請求 請求項の数44 O L (全 42 頁)

(21) 出願番号 特願平7-203259

(22) 出願日 平成7年(1995)8月9日

(71) 出願人 000005223

富士通株式会社

神奈川県川崎市中原区上小田中4丁目1番
1号

(72) 発明者 小原 敦子

神奈川県川崎市中原区上小田中1015番地
富士通株式会社内

(72) 発明者 直井 聡

神奈川県川崎市中原区上小田中1015番地
富士通株式会社内

(74) 代理人 弁理士 大菅 義之 (外1名)

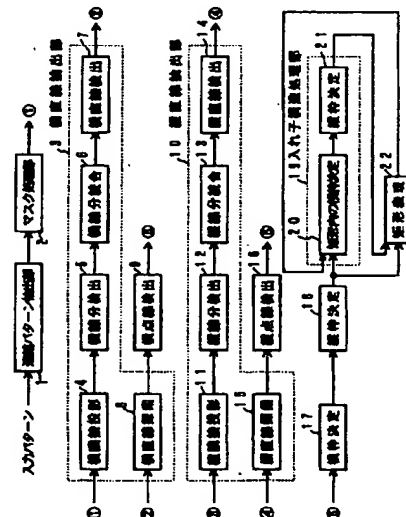
(54) 【発明の名称】 枠抽出装置及び矩形枠抽出装置

(57) 【要約】

【課題】 枠の外形が不規則な矩形構造を有する表形式枠中の枠を正確に抽出し、且つ、文字と枠が接触したり文字が枠からはみ出している場合であっても、該枠を正確に抽出する。

【解決手段】 連結パターン抽出部1により抽出された部分パターンから、横直線抽出部3及び縦直線抽出部10により直線を検出し、横直線抽出部3で抽出された横直線を横枠決定部17により上部から2本づつ組にして横枠とし、縦直線抽出部10で抽出された縦直線の中から縦枠決定部18により縦枠を決定し、2本の横枠と2本の縦枠で囲まれる矩形範囲を、矩形表現部22により枠として抽出する。

本発明の一実施例によるパターン抽出装置の概略構成を示すブロック図



【特許請求の範囲】

【請求項1】 入力画像から、画素の連結情報に基づいて部分パターンを抽出するパターン抽出手段と、該パターン抽出手段により抽出された部分パターンの中から直線を抽出する直線抽出手段と、該直線抽出手段により抽出された直線が、枠を構成する直線であるか否かを判断する判断手段と、該判断手段により得られた枠を構成する直線情報に基づいて、矩形枠を抽出する枠抽出手段と、を備えることを特徴とする枠抽出装置。

【請求項2】 入力画像から、画素の連結情報に基づいて部分パターンを抽出するパターン抽出手段と、該パターン抽出手段により抽出された部分パターンの中から直線を抽出する直線抽出手段と、該直線抽出手段により抽出された直線が、枠を構成する直線であるか否かを判断する判断手段と、該判断手段により得られた枠を構成する直線情報に基づいて、矩形枠を抽出する第一の枠抽出手段と、該第一の枠抽出手段によって抽出された矩形枠内に一定以上の長さを有する横方向の直線が存在する場合、前記矩形領域を新たに表とみなし、該表内に存在する矩形枠を抽出する第二の枠抽出手段と、を備えることを特徴とする枠抽出装置。

【請求項3】 入力画像から、画素の連結情報に基づいて部分パターンを抽出するパターン抽出手段と、該パターン抽出手段により抽出された部分パターンの中から直線を抽出する直線抽出手段と、前記パターン抽出手段により抽出された部分パターンの中から点線を抽出する点線抽出手段と、該直線抽出手段により抽出された直線及び前記点線抽出手段によって抽出された点線が、枠を構成する直線であるか否かを判断する判断手段と、該判断手段により得られた枠を構成する直線情報に基づいて、矩形枠を抽出する枠抽出手段と、を備えることを特徴とする枠抽出装置。

【請求項4】 入力画像から、画素の連結情報に基づいて部分パターンを抽出するパターン抽出手段と、該パターン抽出手段により抽出された部分パターンの中から直線を抽出する直線抽出手段と、前記パターン抽出手段により抽出された部分パターンの中から点線を抽出する点線抽出手段と、該直線抽出手段により抽出された直線及び前記点線抽出手段によって抽出された点線が、枠を構成する直線であるか否かを判断する判断手段と、該判断手段により得られた枠を構成する直線情報に基づいて、矩形枠を抽出する第一の枠抽出手段と、該第一の枠抽出手段によって抽出された矩形枠内に一定以上の長さを有する横方向の直線が存在する場合、前記矩形領域を新たに表とみなし、該表内に存在する矩形枠を抽出する第二の枠抽出手段と、

を備えることを特徴とする枠抽出装置。

【請求項5】 前記直線抽出手段は、ある行又は列の投影値に、周囲の行又は列の投影値を足し合わせた値を、その行又は列の投影値とする隣接投影法を用いて、前記部分パターンを構成する複数の矩形線分を検出する線分検出手段と、該線分検出手段によって得られた複数の線分を統合し、この統合された線分を矩形近似することにより直線を検出する直線検出手段と、

10 を備えることを特徴とする請求項1、2、3または4記載の枠抽出装置。

【請求項6】 前記直線抽出手段は、前記直線検出手段により検出された直線の原画像のパターンの最も細い部分から外側へ探索して、該パターンの両端を検出し、この検出位置を直線の両端の位置として設定する直線探索手段を、さらに、備えることを特徴とする請求項5記載の枠抽出装置。

【請求項7】 該枠抽出手段により抽出された矩形枠の領域の位置情報を、前記線分検出手段によって得られた該矩形枠に統合される前の線分の位置情報を用いて、より細かく設定する矩形枠領域設定手段を、さらに備えることを特徴とする請求項5記載の枠抽出装置。

【請求項8】 前記直線検出手段は、前記隣接投影の範囲を対象画像の解像度に基づいて変化させることを特徴とする請求項5記載の枠抽出装置。

【請求項9】 前記点線抽出手段は、予め設定されているサイズの部分パターンが、所定の方
30 向にある一定の間隔で、一定数以上連続して配置されている画像を検出し、この画像を点線として抽出することを特徴とする請求項3または4記載の枠抽出装置。

【請求項10】 前記点線抽出手段は、上記抽出された点線が存在する領域を、矩形に近似して表現することを特徴とする請求項9記載の枠抽出装置。

【請求項11】 前記点線抽出手段は、上記矩形により表現された点線領域を、さらに細分割して、小矩形で表現することを特徴とする請求項10記載の枠抽出装置。

【請求項12】 前記点線抽出手段は、上記矩形により表現された点線を、直線と同様な位置情報を用いて表現することを特徴とする請求項10又は11記載の枠抽出装置。

【請求項13】 前記判断手段は、前記直線抽出手段によって抽出された横方向の直線の中から、ある一定の長さ以上の直線を抽出して、該直線を横枠線の候補とする横枠線候補抽出手段を、を備えることを特徴とする請求項1、2、3または4記載の枠抽出装置。

50 【請求項14】 前記判断手段は、

前記枠の候補の直線の長さを比較する際に用いる閾値を、前記直線の種類により変化させることを特徴とする請求項 13 記載の枠抽出装置。

【請求項 15】 前記判断手段は、前記横枠線候補抽出手段によって抽出された横枠線の候補の直線を、上部から順番に 2 本ずつ取り出して組とし、この組となった直線を横枠線に決定する横枠決定手段と、該横枠決定手段によって横枠線に決定された 2 本の横線に、上下が共に達している縦方向の 2 本の直線を縦枠線に決定する縦枠決定手段と、を備えることを特徴とする請求項 13 記載の枠抽出装置。

【請求項 16】 前記横枠決定手段は、前記横枠線の候補として抽出した 2 本の組の横方向の直線がほぼ同じ長さである場合、この 2 本の横方向の直線を、横枠の組として登録すること、を特徴とする請求項 15 記載の枠抽出装置。

【請求項 17】 前記横枠決定手段は、前記横枠線の候補として抽出した 2 本の組の横方向の直線のうち、下側の直線の方が長い場合、この 2 本の横方向の直線を、横枠線の組として登録すること、を特徴とする請求項 15 記載の枠抽出装置。

【請求項 18】 前記横枠決定手段は、前記横枠線の候補として抽出した 2 本の横方向の上側の第 1 の直線及び下側の第 2 の直線において、前記第 1 の直線より前記第 2 の直線の方が短く、且つ、前記第 2 の直線の下方に前記第 1 の直線と同じ長さ又は前記第 1 の直線より長い横方向の第 3 の直線が存在しない場合、前記第 1 の直線及び前記第 2 の直線を、横枠線の組として登録すること、を特徴とする請求項 15 記載の枠抽出装置。

【請求項 19】 前記横枠決定手段は、前記横枠線の候補として抽出した 2 本の横方向の上側の第 1 の直線及び下側の第 2 の直線において、前記第 1 の直線より前記第 2 の直線の方が短く、且つ、前記第 2 の直線の下方に前記第 1 の直線と同じ長さ又は前記第 1 の直線より長い横方向の第 3 の直線が存在する場合、前記第 1 の直線及び前記第 3 の直線を、横枠線の組として登録すること、

を特徴とする請求項 15 記載の枠抽出装置。

【請求項 20】 前記横枠決定手段は、前記横枠線の候補として抽出した 2 本の横方向の上側の第 1 の直線及び下側の第 2 の直線において、前記第 1 の直線に対して前記第 2 の直線が右方向又は左方向に短く、且つ、前記第 2 の直線の下方に前記第 1 の直線と同じ長さ又は前記第 1 の直線より長い横方向の第 3 の直線が存在しない場合、前記第 1 の直線及び前記第 2 の直線を、横枠線の組として登録すること、を特徴とする請求項 15 記載の枠抽出装置。

【請求項 21】 前記横枠決定手段は、前記横枠線の候補として抽出した 2 本の横方向の上側の第 1 の直線及び下側の第 2 の直線において、前記第 1 の直線に対して前記第 2 の直線が右方向又は左方向に短く、且つ、前記第 2 の直線の下方に前記第 1 の直線と同じ長さ又は前記第 1 の直線より長い横方向の第 3 の直線が存在する場合、前記第 1 の直線及び前記第 3 の直線を、横枠線の組として登録すること、を特徴とする請求項 15 記載の枠抽出装置。

【請求項 22】 前記横枠決定手段は、前記横枠の候補として抽出した 2 本の横方向の上側の第 1 の直線及び下側の第 2 の直線において縦方向の重なりがない場合、前記第 2 の直線に代えて前記第 2 の直線の下方に存在する横方向の第 3 の直線を、横枠線の候補とすることを特徴とする請求項 15 記載の枠抽出装置。

【請求項 23】 前記横枠決定手段は、前記 2 本の横枠線の内、下側の横枠が右方向へ長い場合又は左方向へ長い場合、前記 2 本の横枠線の両方を横枠線の候補として、再度枠判断処理のために選択可能とすること、を特徴とする請求項 15 記載の枠抽出装置。

【請求項 24】 前記横枠決定手段は、前記 2 本の横枠の長さが異なる場合、長い方の横枠を横枠の候補として再度枠判断処理のために選択可能とすること、を特徴とする請求項 15 記載の枠抽出装置。

【請求項 25】 前記横枠決定手段は、一番下の横枠を決定した後、上部に未処理の横枠の候補が存在している場合、再び部分パターンの上部から未処理の横枠を 2 本ずつ組として枠の候補として抽出すること、を特徴とする請求項 15 記載の枠抽出装置

【請求項 26】 前記縦枠決定手段は、前記横枠決定手段により決定された前記 2 本の横枠に、両端が達している縦線を縦枠として登録すること、を特徴とする請求項 15 記載の枠抽出装置。

【請求項 27】 前記判断手段は、前記直線抽出手段及び前記点線抽出手段によって抽出された横方向の直線または点線の中から、ある一定の長さ以上の線分を抽出して、該線分を横枠線の候補とする横枠線抽出手段を、

備えることを特徴とする請求項 3 または 4 記載の枠抽出装置。

【請求項 28】 前記判断手段は、前記横枠線抽出手段によって抽出された横枠線の候補の直線または点線を、上部から順番に 2 本ずつ取り出して組とし、この組となった直線または点線を横枠線に決定する横枠決定手段と、該横枠決定手段によって横枠線に決定された 2 本の横線に、上下が共に達している縦方向の 2 本の直線を縦枠線に決定する縦枠決定手段と、

を備えることを特徴とする請求項 27 記載の枠抽出装

置。

【請求項 29】 前記横枠決定手段は、前記横枠線の候補として抽出した 2 本の組の横方向の直線がほぼ同じ長さである場合、この 2 本の横方向の直線を、横枠線の組として登録すること、を特徴とする請求項 28 記載の枠抽出装置。

【請求項 30】 前記横枠決定手段は、前記横枠線の候補として抽出した 2 本の組の横方向の直線または点線のうち、下側の直線または点線の方が長い場合、この 2 本の横方向の直線または点線を、横枠線の組として登録すること、を特徴とする請求項 28 記載の枠抽出装置。

【請求項 31】 前記横枠決定手段は、前記横枠線の候補として抽出した 2 本の横方向の上側の第 1 の直線または点線及び下側の第 2 の直線または点線において、前記第 1 の直線または点線より前記第 2 の直線または点線の方が短く、且つ、前記第 2 の直線または点線の下方に前記第 1 の直線または点線と同じ長さ又は前記第 1 の直線または点線より長い横方向の第 3 の直線または点線が存在しない場合、前記第 1 の直線または点線及び前記第 2 の直線または点線を、横枠線の組として登録すること、を特徴とする請求項 28 記載の枠抽出装置。

【請求項 32】 前記横枠決定手段は、前記横枠線の候補として抽出した 2 本の横方向の上側の第 1 の直線または点線及び下側の第 2 の直線または点線において、前記第 1 の直線または点線より前記第 2 の直線または点線の方が短く、且つ、前記第 2 の直線または点線の下方に前記第 1 の直線または点線と同じ長さ、又は前記第 1 の直線または点線より長い横方向の第 3 の直線または点線が存在する場合、前記第 1 の直線または点線及び前記第 3 の直線または点線を、横枠線の組として登録すること、を特徴とする請求項 28 記載の枠抽出装置。

【請求項 33】 前記横枠決定手段は、前記横枠線の候補として抽出した 2 本の横方向の上側の第 1 の直線または点線及び下側の第 2 の直線または点線において、前記第 1 の直線または点線に対して前記第 2 の直線または点線が右方向又は左方向に短く、且つ、前記第 2 の直線または点線の下方に前記第 1 の直線または点線と同じ長さ又は前記第 1 の直線または点線より長い横方向の第 3 の直線または点線が存在しない場合、前記第 1 の直線または点線及び前記第 2 の直線または点線を、横枠線の組として登録すること、を特徴とする請求項 28 記載の枠抽出装置。

【請求項 34】 前記横枠決定手段は、前記横枠線の候補として抽出した 2 本の横方向の上側の第 1 の直線または点線及び下側の第 2 の直線または点線において、前記第 1 の直線または点線に対して前記第 2 の直線または点線が右方向又は左方向に短く、且つ、前

記第 2 の直線または点線の下方に前記第 1 の直線または点線と同じ長さ又は前記第 1 の直線または点線より長い横方向の第 3 の直線または点線が存在する場合、前記第 1 の直線または点線及び前記第 3 の直線または点線を、横枠線の組として登録すること、を特徴とする請求項 28 記載の枠抽出装置。

【請求項 35】 前記横枠決定手段は、前記横枠の候補として抽出した 2 本の横方向の上側の第 1 の直線または点線及び下側の第 2 の直線または点線において縦方向の重なりがない場合、前記第 2 の直線または点線に代えて前記第 2 の直線または点線の下方に存在する横方向の第 3 の直線または点線を、横枠線の候補とすること、を特徴とする請求項 28 記載の枠抽出装置。

【請求項 36】 前記横枠決定手段は、前記 2 本の横枠線の内、下側の横枠が右方向へ長い又は左方向へ長い場合、前記 2 本の横枠線の両方を横枠線の候補として、再度枠判断処理のために選択可能とすることを特徴とする請求項 28 記載の枠抽出装置。

【請求項 37】 前記横枠決定手段は、前記 2 本の横枠の長さが異なる場合、長い方の横枠を横枠の候補として再度枠判断処理のために選択可能とすること、を特徴とする請求項 28 記載の枠抽出装置。

【請求項 38】 前記横枠決定手段は、一番下の横枠を決定した後、上部に未処理の横枠の候補が存在している場合、再び部分パターンの上部から未処理の横枠を 2 本づつ組として枠の候補を抽出することを特徴とする請求項 28 記載の枠抽出装置。

【請求項 39】 前記縦枠決定手段は、前記横枠決定手段により決定された前記 2 本の横枠に、両端が達している縦線を縦枠として登録すること、を特徴とする請求項 28 記載の枠抽出装置。

【請求項 40】 前記点線抽出手段は、点線を抽出する際、構成する点線間の間隔に応じて、原画像の両端の点の位置を外延させて点線を表現する矩形の範囲を決定すること、を特徴とする請求項 3 または 4 記載の枠抽出装置。

【請求項 41】 入力画像から、画素の連結情報に基づいて部分パターンを抽出するパターン抽出手段と、前記パターン抽出手段により抽出された部分パターンの中から点線を抽出する点線抽出手段と、該点線抽出手段によって抽出された点線が、枠を構成する直線であるか否かを判断する判断手段と、該判断手段により得られた枠を構成する直線情報に基づいて、矩形枠を抽出する枠抽出手段と、を備えることを特徴とする枠抽出装置。

【請求項 42】 入力画像から、画素が連結している部分パターンを抽出する連結パターン抽出手段と、該連結パターン抽出手段により抽出された部分パターンの中から、所定の長さ以上の直線を抽出する直線抽出手

段と、

該直線抽出手段により抽出された直線のうち、横直線を上から順番に2本づつ組にして横枠線とする横枠線決定手段と、

該直線抽出手段により抽出された直線のうち、上端及び下端が前記横枠線に達している縦直線を縦枠線とする縦枠線決定手段と、

前記横枠線決定手段によって組とされた2本の横枠線と、前記縦枠線決定手段により決定された2本の縦枠線とで、囲まれた矩形範囲を抽出する矩形抽出手段と、を備えることを特徴とする矩形抽出装置。

【請求項43】 パターンの連結情報に基づいて部分パターンを抽出するパターン抽出手段と、前記パターン抽出手段により抽出された部分パターンに対して直線を抽出する直線抽出手段と、前記パターン抽出手段により抽出された部分パターンに対して点線を抽出する点線抽出手段と、前記直線抽出手段及び前記点線抽出手段により抽出された直線及び点線が枠であるかどうかを判断する判断手段と、

前記枠の情報に基づいて文字領域を抽出する文字領域抽出手段とを備えることを特徴とする文字切り出し方式。

【請求項44】 入力されたパターンから連結した部分パターンを抽出する連結パターン抽出手段と、前記連結パターン抽出手段により抽出された部分パターンから所定の長さ以上の直線を抽出する直線抽出手段と、

前記直線抽出手段により抽出された直線のうち、横直線を上から順番に2本づつ組にして横枠線とする横枠線決定手段と、

前記直線抽出手段により抽出された直線のうち、上端及び下端が前記横枠線に達している縦直線を縦枠線とする縦枠線決定手段と、

前記横枠線の2本と前記縦枠線の2本とで囲まれた矩形範囲を抽出する矩形表現手段とを備えることを特徴とするパターン抽出装置。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】本発明は、枠抽出装置及び矩形抽出装置に係わり、特に、手書き用文字認識装置、印刷文字認識装置、図面認識装置等において、文字枠などのような枠や矩形等を抽出する枠抽出装置及び矩形抽出装置に関する

【0002】

【従来の技術】近年、手書き用の入力周辺機器としてOCRなどの手書き文字認識装置の需要が増加している。そして、手書き文字認識装置において個々の文字の高い認識率を実現するためには、認識の前処理である文字の切り出し処理を正確に行うことが重要になる。

【0003】これまで、帳票などの予め文字を書く位置

を指定した文書（ドロップアウトカラーでなく、黒枠などの罫線と文字とが同じ色で、且つ、同じ濃度で書かれたもの）では、指定した範囲内に文字が正確に書かれた場合は、高い確率で文字を正確に自動認識することができた。しかしながら、文字が指定した範囲を越えて指定範囲を示す枠や罫線に文字が接触したり、枠や罫線から文字がはみ出した場合は、著しく文字の認識率が低下する問題が生じていた。すなわち、現在使用されている文字認識装置では、枠を除去する際、文字枠のわずかな傾きや凹凸に対応することができず、文字枠の位置や線幅の変動により本来の文字部分の所が欠けてしまったり、枠が残存したりする。

【0004】これらの問題点に対応するため、本出願人は、以前、例えば、特開平6-309498号公報や特開平7-28937号公報に記載されているように、図42(a)に示す一文字枠、図42(b)に示す横一行ブロック枠、図42(c)に示すフリーフォーマット枠、図42(d)(e)に示す横枠線が規則的に配置されている表形式枠について、枠の位置やサイズ等のフォーマット情報の入力が必要とせずに、枠の抽出や除去を行うことが可能な技術を提案した。

【0005】以下、これらの技術の概要について説明する。まず、一文字枠の抽出や除去を行う方法では、パターンの線幅の検出を行い、そのパターンの線幅が太い場合、図43(a)に示すように、横方向及び縦方向にパターンの投影をとる。すなわち、パターンを構成する各画素の濃度を横方向及び縦方向に加算し、加算して得られた投影値が一定閾値以上である部分を枠とみなす。

【0006】また、パターンの線幅が細い場合、図43(b)に示すように、パターンの端から、nラインランレングスという8連結で接続するランレングスの手法により、規定のnラインの間で探索を行い、線の長さを算出する。そして、線の長さが一定閾値以上であれば、その部分を枠とみなす。

【0007】以上の方法で多数の枠線の候補が出た場合には、一番外側の2本を枠とする。また、横一行ブロック枠の抽出や除去を行う方法では、図44(a)の原画像に対し、マスク処理による細線化処理を行う。すなわち、図44(b)に示すように、原画像におけるパターンの縦横線分を細線化し、文字と枠の太さの差をなくす。そして、直線抽出処理により、比較的長い直線の抽出を行う。この直線抽出処理は、隣接投影法により行われ、ある画素の投影値にその周囲の画素の投影値を足しあわせた結果を最終的な投影値とし、その投影値が一定閾値以上である部分を直線とみなすものである。

【0008】次に、横枠決定処理を行う。この横枠決定処理では、上記直線抽出処理で抽出された直線を横方向（水平方向）に探索し、途中で途切れていないかどうかを検査する。そして、図44(c)に示すように、途切れない直線を統合して長い直線を抽出する。そして、

横枠線として統合された複数の横線の中から一番長いものを、横枠線として選択する。

【0009】次に、縦枠決定処理を行う。この縦枠決定処理では、図44(e)に示すように、線間隔が等間隔になるような縦線を枠線として選択する。これは、図44(c)に示すように、文字が横枠線にかかっている場合、縦線の長さだけでは枠線か文字線かを判断できないためである。縦線の間隔の決定は、線間隔のヒストグラムを求め、最も頻度の高い線間隔を枠の間隔とすることにより行う。例えば、図44(d)において、間隔xで並んでいる縦線を縦枠とみなし、間隔y、zの縦線を縦枠から除外する。尚、全ての線間隔が異なっていた場合にはフリーフォーマット枠とみなし、左右両端の縦線のみを縦枠とする。

【0010】次に、枠除去処理を行う。この枠除去処理では、図44(f)に示すように、上述のようにして決定された縦と横の枠線の交点を検出して、上記横枠と左右両端の縦枠で決定される矩形枠を一文字枠の範囲に分割する。そして、それぞれの分割された一文字枠の範囲内で、公知の一文字枠の除去処理と同様な処理により枠を除去する。

【0011】また、上述した縦枠決定処理において、縦直線が左端と右端のみに存在した場合、又は、縦直線どうしの線間隔が全て異なっていた場合には、上記横枠と縦枠で決定される矩形をフリーフォーマット枠と判定する。

【0012】次に、図45に示すように、このフリーフォーマット枠をある一定長さで分割する。そして、上述した一文字枠の抽出や除去の処理のときと同様に、その分割した範囲内で横方向及び縦方向にパターンの投影をとることにより、枠部分の決定を行う。また、それぞれの分割部分がフリーフォーマット枠の中のどこに位置するかにより、消去する枠の本数と位置とを変化させる。例えば、その分割部分が左端なら、左、上、下の枠を消去し、その分割部分が右端の場合は右、上、下の枠を消去し、その分割部分が真ん中の場合は上、下の枠のみ消去を行う。

【0013】また、表形式ブロック枠の抽出や除去を行う方法では、まず、横線抽出処理を行う。この横線抽出処理では、上述した横一行ブロック枠の抽出や除去を行う方法で示したように、途切れのない直線を統合して得られた横線を連結パターンの外接矩形サイズと比較し、その横線がある一定以上の長さであった場合、その横線を横枠とする。ここで、横枠が2本得られた場合には、対象とするパターンは横1行のブロック枠又はフリーフォーマット枠であり、横枠が3本以上得られた場合には、対象とするパターンは表形式ブロック枠と判定する。

【0014】表形式ブロック枠の場合、横枠の中で隣り合う2本を選択し、それらの2本の横枠を上下の横枠と

する1行のブロック枠が縦に重なっているものと考え、この結果、表形式ブロック枠を横1行ブロック枠と同様に扱うことが可能となる。例えば、図46(a)

(c)(e)に示す4本の横枠を有する表形式ブロック枠は、図46(b)(d)(f)に示す一行目～三行目の横一行ブロック枠に分割される。

【0015】次に、縦枠検出処理を行う。この縦枠検出処理は、隣り合う2本の横枠には含まれた領域にある縦直線の中から、縦枠の検出を行うもので、検出された縦直線を探索することによって、上下の横枠に達しているか、また途中で途切れていないかを調べる。そして、探索の結果、上下が共に横枠に達して、かつ途中で途切れていない縦直線を縦枠の候補とする。

【0016】次に、この縦枠検出処理で得られた縦枠の候補の中から縦枠を決定する縦枠決定処理を行う。この縦枠決定処理は、縦線の並び方により以下のように処理が異なる。

【0017】まず、図46(a)に示すように、縦線が規則的に並んでいる表である場合、縦枠の候補に残った縦直線どうしの間隔を算出する。そして、該線間隔とその出現頻度をヒストグラムに示し、著しく他と異なる間隔を有する縦直線を枠の候補から除外する。そして、枠候補から除外されなかった縦直線を縦枠と決定する。

【0018】一方、図46(c)に示すように、縦線の不規則に並んでいる表である場合、縦枠の候補に残った縦線どうしの距離が文字サイズと比較して極端に小さい場合、該縦線を枠の候補から除外する。そして、枠の候補から除外されなかった縦線を全て縦枠であると判定する。

【0019】また、図46(e)に示すように縦枠が左端と右端の2本しかないパターンの場合、又は、全ての縦枠の線間隔が異なっているパターンの場合、これらのパターンをフリーフォーマット枠と判定し、左端と右端の2本の縦直線を縦枠と決定する。

【0020】

【発明が解決しようとする課題】従来の文字認識装置において処理可能な帳票は、図42に示すような一文字枠、ブロック枠（横一行枠またはフリーフォーマット枠）、または枠の外形が矩形状となっており横枠線が規則的に配置されている構造を持つ表を有するものであった。また、帳票の枠抽出処理で対象としていた罫線は実線のみであり、点線を含んだ帳票に関しては処理の対象外であった。

【0021】しかし、処理可能な帳票、上記のような一文字枠、ブロック枠、枠の外形が矩形状となっており横枠線が規則的に配置されている構造を持つ表を有する帳票のみでは、文字認識に利用可能な帳票の種類が限定されてしまい、様々な形状の枠や表を有する一般帳票の対応には不十分である。

【0022】すなわち、例えば、図47に示すように、

外形が矩形となっていない表を有する帳票、また表の一部分にさらに細かい表構造(図47の網かけ部分)を有する帳票のなどに対しては、従来の文字認識装置では対応することができなかった。さらに、点線と実線とが混在している帳票などに対しても、対応することができなかった。本発明の目的は、枠の外形が矩形状となっていない表の枠を正確に抽出できるようにすることである。

【0023】また、点線と実線とが混在している表の枠を正確に抽出できるようにすることである。また、さらに、入れ子構造を有する表の枠を正確に抽出できるようにすることである。また、上記枠抽出情報を基に文字領域を正確に抽出することができるようにすることである。

【0024】

【課題を解決するための手段】上述した課題を解決するために、請求項1の発明によれば、入力画像から、画素の連結情報に基づいて部分パターンを抽出するパターン抽出手段と、該パターン抽出手段により抽出された部分パターンの中から直線を抽出する直線抽出手段と、該直線抽出手段により抽出された直線が、枠を構成する直線であるか否かを判断する判断手段と、該判断手段により得られた枠を構成する直線情報に基づいて、矩形枠を抽出する枠抽出手段とを備える。このことにより、様々な構造の表形式枠を抽出することができる。

【0025】また、請求項2の発明によれば、入力画像から、画素の連結情報に基づいて部分パターンを抽出するパターン抽出手段と、該パターン抽出手段により抽出された部分パターンの中から直線を抽出する直線抽出手段と、該直線抽出手段により抽出された直線が、枠を構成する直線であるか否かを判断する判断手段と、該判断手段により得られた枠を構成する直線情報に基づいて、矩形枠を抽出する第一の枠抽出手段と、該第一の枠抽出手段によって抽出された矩形枠内に一定以上の長さを有する横方向の直線が存在する場合、前記矩形領域を新たに表とみなし、該表内に存在する矩形枠を抽出する第二の枠抽出手段とを備える。このことにより、一定以上の長さを有する横方向の直線又は点線が抽出した矩形領域内に存在する場合、その矩形領域を新たに表とみなすことができ、複雑な構造を有する表であっても、枠の抽出や文字領域の抽出が可能となる。

【0026】また、請求項3の発明によれば、入力画像から、画素の連結情報に基づいて部分パターンを抽出するパターン抽出手段と、該パターン抽出手段により抽出された部分パターンの中から直線を抽出する直線抽出手段と、前記パターン抽出手段により抽出された部分パターンの中から点線を抽出する点線抽出手段と、該直線抽出手段により抽出された直線及び前記点線抽出手段によって抽出された点線が、枠を構成する直線であるか否かを判断する判断手段と、該判断手段により得られた枠を構成する直線情報に基づいて、矩形枠を抽出する枠抽出

手段とを備える。このことにより、点線と実線とが混在している表の枠を正確に抽出できる。

【0027】また、請求項4の発明によれば、入力画像から、画素の連結情報に基づいて部分パターンを抽出するパターン抽出手段と、該パターン抽出手段により抽出された部分パターンの中から直線を抽出する直線抽出手段と、前記パターン抽出手段により抽出された部分パターンの中から点線を抽出する点線抽出手段と、該直線抽出手段により抽出された直線及び前記点線抽出手段によって抽出された点線が、枠を構成する直線であるか否かを判断する判断手段と、該判断手段により得られた枠を構成する直線情報に基づいて、矩形枠を抽出する第一の枠抽出手段と、該第一の枠抽出手段によって抽出された矩形枠内に一定以上の長さを有する横方向の直線が存在する場合、前記矩形領域を新たに表とみなし、該表内に存在する矩形枠を抽出する第二の枠抽出手段とを備える。このことにより、点線と実線とが混在している矩形領域を新たに表とみなすことができ、複雑な構造を有し、且つ、点線と実線とが混在する表であっても、枠の抽出や文字領域の抽出が可能となる。

【0028】また、請求項5の発明によれば、前記直線抽出手段は、ある行又は列の投影値に、周囲の行又は列の投影値を足し合わせた値を、その行又は列の投影値とする隣接投影法を用いて、前記部分パターンを構成する複数の矩形線分を検出する線分検出手段と、該線分検出手段によって得られた複数の線分を統合し、この統合された線分を矩形近似することにより直線を検出する直線検出手段とを備える。このことにより、周囲の行又は列の投影値を足し合わせた投影値により矩形近似して直線を検出することができ、斜めに傾いた直線を正確に検出することができる。

【0029】また、請求項6の発明によれば、前記直線抽出手段は、前記直線検出手段により検出された直線の原画像のパターンの最も細い部分から外側へ探索して、該パターンの両端を検出し、この検出位置を直線の両端の位置として設定する直線探索手段を備える。このことにより、直線の両端の位置を正確に検出することが可能となる。

【0030】また、請求項7の発明によれば、該枠抽出手段により抽出された矩形枠の領域の位置情報を、前記線分検出手段によって得られた該矩形枠に統合される前の線分の位置情報を用いて、より細かく設定する矩形枠領域設定手段を備える。このことにより、直線の位置をより正確に検出することが可能となる。

【0031】また、請求項8の発明によれば、前記直線抽出手段は、前記隣接投影の範囲を対象画像の解像度に基づいて変化させる。このことにより、直線の抽出をより正確に行うことが可能となる。

【0032】また、請求項9の発明によれば、前記点線抽出手段は、予め設定されているサイズの部分パターン

が、所定の方向にある一定の間隔で、一定数以上連続して配置されている画像を検出し、この画像を点線として抽出する。このことにより、部分パターンのサイズ及び部分パターン間の距離により点線を抽出することができ、処理時間の短縮を図ることが可能となる。

【0033】また、請求項10の発明によれば、前記点線抽出手段は、上記抽出された点線が存在する領域を、矩形に近似して表現する。このことにより、点線を矩形で近似された実線と同様に扱うことが可能となる。

【0034】また、請求項11の発明によれば、前記点線抽出手段は、上記矩形により表現された点線領域を、さらに細分割して、小矩形で表現する。このことにより、傾きがあっても正確に点線の存在する位置を表現することが可能となる。

【0035】また、請求項12の発明によれば、前記点線抽出手段は、上記矩形により表現された点線を、直線と同様な位置情報を用いて表現する。このことにより、点線を実線と同様に扱うことが可能となる。

【0036】また、請求項13の発明によれば、前記判断手段は、前記直線抽出手段によって抽出された横方向の直線の中から、ある一定の長さ以上の直線を抽出して、該直線を横枠線の候補とする横枠線候補抽出手段を備える。このことにより、枠の大きさと比較すると非常に短い横線であっても、枠の候補として抽出することが可能である。

【0037】また、請求項14の発明によれば、前記判断手段は、前記枠の候補の直線の長さを比較する際に用いる閾値を、前記直線の種類により変化させる。このことにより、実線と点線の混在している場合であっても正確に枠を決定することが可能となる。

【0038】また、請求項15の発明によれば、前記判断手段は、前記横枠線候補抽出手段によって抽出された横枠線の候補の直線を、上部から順番に2本ずつ取り出して組とし、この組となった直線を横枠線に決定する横枠決定手段と、該横枠決定手段によって横枠線に決定された2本の横線に、上下が共に達している縦方向の2本の直線を縦枠線に決定する縦枠決定手段とを備える。このことにより、長さが異なる横直線を横枠の組として処理することができ、枠の外形が矩形でない場合にも枠の抽出が可能となる。

【0039】また、請求項16の発明によれば、前記横枠決定手段は、前記横枠線の候補として抽出した2本の組の横方向の直線がほぼ同じ長さである場合、この2本の横方向の直線を、横枠の組として登録する。このことにより、枠の外形が矩形の場合の枠の抽出が可能となる。

【0040】また、請求項17の発明によれば、前記横枠決定手段は、前記横枠線の候補として抽出した2本の組の横方向の直線のうち、下側の直線の方が長い場合、この2本の横方向の直線を、横枠線の組として登録す

る。このことにより、枠の外形が矩形でない場合にも枠の抽出が可能となる。

【0041】また、請求項18の発明によれば、前記横枠決定手段は、前記横枠線の候補として抽出した2本の横方向の上側の第1の直線及び下側の第2の直線において、前記第1の直線より前記第2の直線の方が短く、且つ、前記第2の直線の下方に前記第1の直線と同じ長さ又は前記第1の直線より長い横方向の第3の直線が存在しない場合、前記第1の直線及び前記第2の直線を、横枠線の組として登録する。このことにより、枠の外形が矩形でない場合にも枠の抽出が可能となる。

【0042】また、請求項19の発明によれば、前記横枠決定手段は、前記横枠線の候補として抽出した2本の横方向の上側の第1の直線及び下側の第2の直線において、前記第1の直線より前記第2の直線の方が短く、且つ、前記第2の直線の下方に前記第1の直線と同じ長さ又は前記第1の直線より長い横方向の第3の直線が存在する場合、前記第1の直線及び前記第3の直線を、横枠線の組として登録する。このことにより、枠の外形が矩形でない場合にも枠の抽出が可能となる。

【0043】また、請求項20の発明によれば、前記横枠決定手段は、前記横枠線の候補として抽出した2本の横方向の上側の第1の直線及び下側の第2の直線において、前記第1の直線に対して前記第2の直線が右方向又は左方向に短く、且つ、前記第2の直線の下方に前記第1の直線と同じ長さ又は前記第1の直線より長い横方向の第3の直線が存在しない場合、前記第1の直線及び前記第2の直線を、横枠線の組として登録する。このことにより、枠の外形が矩形でなく、且つ、上側の直線と下側の直線とが横方向にずれて配置されている場合にも枠の抽出が可能となる。

【0044】また、請求項21の発明によれば、前記横枠決定手段は、前記横枠線の候補として抽出した2本の横方向の上側の第1の直線及び下側の第2の直線において、前記第1の直線に対して前記第2の直線が右方向又は左方向に短く、且つ、前記第2の直線の下方に前記第1の直線と同じ長さ又は前記第1の直線より長い横方向の第3の直線が存在する場合、前記第1の直線及び前記第3の直線を、横枠線の組として登録する。このことにより、枠の外形が矩形でなく、且つ、上側の直線と下側の直線とが横方向にずれて配置されている場合にも枠の抽出が可能となる。

【0045】また、請求項22の発明によれば、前記横枠決定手段は、前記横枠線の候補として抽出した2本の横方向の上側の第1の直線及び下側の第2の直線において縦方向の重なりがない場合、前記第2の直線に代えて前記第2の直線の下方に存在する横方向の第3の直線を、横枠線の候補とする。このことにより、枠の外形が矩形でなく、且つ、上側の直線と下側の直線とが縦方向にずれて配置されている場合にも枠の抽出が可能となる。

【0046】また、請求項23の発明によれば、前記横枠決定手段は、前記2本の横枠線の内、下側の横枠が右方向へ長い場合又は左方向へ長い場合、前記2本の横枠線の両方を横枠線の候補として、再度枠判断処理のために選択可能とする。このことにより、枠の外形が矩形でなく、且つ、上側の直線と下側の直線とが横方向にずれて配置されている場合においても、未処理の横枠の組を作ることが可能となる。

【0047】また、請求項24の発明によれば、前記横枠決定手段は、前記2本の横枠の長さが異なる場合、長い方の横枠を横枠の候補として再度枠判断処理のために選択可能とする。このことにより、上から順に横枠の組を作った場合において、未処理として残った横線を横枠の組とすることが可能となる。

【0048】また、請求項25の発明によれば、前記横枠決定手段は、一番下の横枠を決定した後、上部に未処理の横枠の候補が存在している場合、再び部分パターンの上部から未処理の横枠を2本ずつ組として枠の候補として抽出する。このことにより、上から順に横枠の組を作った場合において、未処理として残った横線を横枠の組とすることが可能となる。

【0049】また、請求項26の発明によれば、前記縦枠決定手段は、前記横枠決定手段により決定された前記2本の横枠に、両端が達している縦線を縦枠として登録する。このことにより、縦枠を抽出する場合において、文字と枠とを正確に区別することができる。

【0050】また、請求項27の発明によれば、前記判断手段は、前記直線抽出手段及び前記点線抽出手段によって抽出された横方向の直線または点線の中から、ある一定の長さ以上の線分を抽出して、該線分を横枠線の候補とする横枠線抽出手段を備える。このことにより、横枠を抽出する場合において、文字と枠とを正確に区別することができる。

【0051】また、請求項28の発明によれば、前記判断手段は、前記横枠線抽出手段によって抽出された横枠線の候補の直線または点線を、上部から順番に2本ずつ取り出して組とし、この組となった直線または点線を横枠線に決定する横枠決定手段と、該横枠決定手段によって横枠線に決定された2本の横線に、上下が共に達している縦方向の2本の直線を縦枠線に決定する縦枠決定手段とを備える。このことにより、長さが異なる横直線または横点線を横枠の組として処理することができ、枠の外形が矩形でなく、且つ、点線が混在している場合にも枠の抽出が可能となる。

【0052】また、請求項29の発明によれば、前記横枠決定手段は、前記横枠線の候補として抽出した2本の組の横方向の直線がほぼ同じ長さである場合、この2本の横方向の直線を、横枠線の組として登録する。このことにより、枠の外形が矩形で点線が混在している場合の枠の抽出が可能となる。

【0053】また、請求項30の発明によれば、前記横枠決定手段は、前記横枠線の候補として抽出した2本の組の横方向の直線または点線のうち、下側の直線または点線の方が長い場合、この2本の横方向の直線または点線を、横枠線の組として登録する。このことにより、枠の外形が矩形でなく、且つ、点線が混在している場合にも枠の抽出が可能となる。

【0054】また、請求項31の発明によれば、前記横枠決定手段は、前記横枠線の候補として抽出した2本の横方向の上側の第1の直線または点線及び下側の第2の直線または点線において、前記第1の直線または点線より前記第2の直線または点線の方が短く、且つ、前記第2の直線または点線の下方に前記第1の直線または点線と同じ長さ又は前記第1の直線または点線より長い横方向の第3の直線または点線が存在しない場合、前記第1の直線または点線及び前記第2の直線または点線を、横枠線の組として登録する。このことにより、枠の外形が矩形でなく、且つ、点線が混在している場合にも枠の抽出が可能となる。

【0055】また、請求項32の発明によれば、前記横枠決定手段は、前記横枠線の候補として抽出した2本の横方向の上側の第1の直線または点線及び下側の第2の直線または点線において、前記第1の直線または点線より前記第2の直線または点線の方が短く、且つ、前記第2の直線または点線の下方に前記第1の直線または点線と同じ長さ、又は前記第1の直線または点線より長い横方向の第3の直線または点線が存在する場合、前記第1の直線または点線及び前記第3の直線または点線を、横枠線の組として登録する。このことにより、枠の外形が矩形でなく、且つ、点線が混在している場合にも枠の抽出が可能となる。

【0056】また、請求項33の発明によれば、前記横枠決定手段は、前記横枠線の候補として抽出した2本の横方向の上側の第1の直線または点線及び下側の第2の直線または点線において、前記第1の直線または点線に対して前記第2の直線または点線が右方向又は左方向に短く、且つ、前記第2の直線または点線の下方に前記第1の直線または点線と同じ長さ又は前記第1の直線または点線より長い横方向の第3の直線または点線が存在しない場合、前記第1の直線または点線及び前記第2の直線または点線を、横枠線の組として登録する。このことにより、枠の外形が矩形でなく、且つ、上側の直線または点線と下側の直線または点線とが横方向にずれて配置されている場合にも枠の抽出が可能となる。

【0057】また、請求項34の発明によれば、前記横枠決定手段は、前記横枠線の候補として抽出した2本の横方向の上側の第1の直線または点線及び下側の第2の直線または点線において、前記第1の直線または点線に対して前記第2の直線または点線が右方向又は左方向に短く、且つ、前記第2の直線または点線の下方に前記第

1の直線または点線と同じ長さ又は前記第1の直線または点線より長い横方向の第3の直線または点線が存在する場合、前記第1の直線または点線及び前記第3の直線または点線を、横枠線の組として登録する。このことにより、枠の外形が矩形でなく、且つ、上側の直線または点線と下側の直線または点線とが横方向にずれて配置されている場合にも枠の抽出が可能となる。

【0058】また、請求項35の発明によれば、前記横枠決定手段は、前記横枠の候補として抽出した2本の横方向の上側の第1の直線または点線及び下側の第2の直線または点線において縦方向の重なりがない場合、前記第2の直線または点線に代えて前記第2の直線または点線の下方に存在する横方向の第3の直線または点線を、横枠線の候補とする。このことにより、枠の外形が矩形でなく、且つ、上側の直線または点線と下側の直線または点線とが縦方向にずれて配置されている場合にも枠の抽出が可能となる。

【0059】また、請求項36の発明によれば、前記横枠決定手段は、前記2本の横枠線の内、下側の横枠が右方向へ長い又は左方向へ長い場合、前記2本の横枠線の両方を横枠線の候補として、再度枠判断処理のために選択可能とする。このことにより、枠の外形が矩形でなく、且つ、上側の直線または点線と下側の直線または点線とが横方向にずれて配置されている場合においても、未処理の横枠の組を作ることが可能となる。

【0060】また、請求項37の発明によれば、前記横枠決定手段は、前記2本の横枠の長さが異なる場合、長い方の横枠を横枠の候補として再度枠判断処理のために選択可能とする。このことにより、このことにより、上から順に横枠の組を作った場合において、未処理として残った横線または横点線を横枠の組とすることが可能となる。

【0061】また、請求項38の発明によれば、前記横枠決定手段は、一番下の横枠を決定した後、上部に未処理の横枠の候補が存在している場合、再び部分パターンの上部から未処理の横枠を2本ずつ組として枠の候補を抽出する。このことにより、このことにより、上から順に横枠の組を作った場合において、未処理として残った横線または横点線を横枠の組とすることが可能となる。

【0062】また、請求項39の発明によれば、前記縦枠決定手段は、前記横枠決定手段により決定された前記2本の横枠に、両端が達している縦線を縦枠として登録する。このことにより、直線と点線とが混在する縦枠を抽出する場合において、文字と枠とを正確に区別することができる。

【0063】また、請求項40の発明によれば、前記点線抽出手段は、点線を抽出する際、構成する点線間の間隔に応じて、原画像の両端の点の位置を外延させて点線を表現する矩形の範囲を決定する。このことにより、よ

り正確に点線の位置を記述することが可能となる。

【0064】また、請求項41の発明によれば、入力画像から、画素の連結情報に基づいて部分パターンを抽出するパターン抽出手段と、前記パターン抽出手段により抽出された部分パターンの中から点線を抽出する点線抽出手段と、該点線抽出手段によって抽出された点線が、枠を構成する直線であるか否かを判断する判断手段と、該判断手段により得られた枠を構成する直線情報に基づいて、矩形枠を抽出する枠抽出手段とを備える。このことにより、横点線の長さが異なっている場合においても、横点線の組を作ることができ、枠の外形が矩形状となっていない点線で構成された表の枠を正確に抽出できる。

【0065】また、請求項42の発明によれば、入力画像から、画素が連結している部分パターンを抽出する連結パターン抽出手段と、該連結パターン抽出手段により抽出された部分パターンの中から、所定の長さ以上の直線を抽出する直線抽出手段と、該直線抽出手段により抽出された直線のうち、横直線を上から順番に2本ずつ組にして横枠線とする横枠線決定手段と、該直線抽出手段により抽出された直線のうち、上端及び下端が前記横枠線に達している縦直線を縦枠線とする縦枠線決定手段と、前記横枠線決定手段によって組とされた2本の横枠線と、前記縦枠線決定手段により決定された2本の縦枠線とで、囲まれた矩形範囲を抽出する矩形抽出手段とを備える。このことにより、横直線の長さが異なっている場合においても、横直線の組を作ることができ、枠の外形が矩形状となっていない表の枠を正確に抽出できる。

【0066】また、請求項43の発明によれば、パターン連結情報に基づいて部分パターンを抽出するパターン抽出手段と、前記パターン抽出手段により抽出された部分パターンに対して直線を抽出する直線抽出手段と、前記パターン抽出手段により抽出された部分パターンに対して点線を抽出する点線抽出手段と、前記直線抽出手段及び前記点線抽出手段により抽出された直線及び点線が枠であるかどうかを判断する判断手段と、前記枠の情報に基づいて文字領域を抽出する文字領域抽出手段とを備える。このことにより、連結した部分パターンから直線又は点線を抽出し、抽出した直線又は点線が枠あるかどうかを判断し、その枠の情報に基づいて文字領域を抽出することができ、様々な構造の表形式枠における文字領域を正確に抽出することが可能となる。

【0067】また、請求項44の発明によれば、入力されたパターンから連結した部分パターンを抽出する連結パターン抽出手段と、前記連結パターン抽出手段により抽出された部分パターンから所定の長さ以上の直線を抽出する直線抽出手段と、前記直線抽出手段により抽出された直線のうち、横直線を上から順番に2本ずつ組にして横枠とする横枠決定手段と、前記直線抽出手段により抽出された直線のうち、上端及び下端が前記横枠に達し

ている縦直線を縦枠とする縦枠決定手段と、前記横枠の2本と前記縦枠の2本とで囲まれた矩形範囲を抽出する矩形表現手段とを備える。このことにより、横直線の長さが異なっている場合においても、横直線の組を作ることができ、枠の外形が矩形状となっていない表の枠を正確に抽出できる。

【0068】

【発明の実施の形態】以下、本発明の実施例を、図面を参照しながら詳細に説明する。図1は、本発明の一実施例であるパターン抽出装置の構成を示すブロック図である。

【0069】このパターン抽出装置は、文字の書き込み領域が規則的な構造を有する表だけではなく、不規則な構造を有する表においても、枠を正確に抽出できるようにしたものである。また、実線と点線の罫線が混在する表や、入れ子構造を有する表にも適用可能である。

【0070】尚、以下の説明では、枠が単数または複数個あり、また、サイズや位置や傾きが分からない枠に対して文字が書かれている（この場合、文字が枠と接触したり、枠からはみ出しているもよい）帳票から枠を抽出する場合について説明する。また、対象となる入力パターンは、極端な傾きや回転の補正が行われ、さらに、かすれの穴埋めなどの前処理が施された後の2値画像である。

【0071】図1に示すように、本実施例のパターン抽出装置は、連結パターン抽出部1、マスク処理部2、横直線抽出部3、横点線検出部9、縦直線抽出部10、縦点線検出部16、横枠決定部17、縦枠決定部18、入れ子構造処理部19及び矩形表現部22を備えている。

【0072】横直線抽出部3は、横隣接投影部4、横線分検出部5、横線分統合部6、横直線検出部7及び横直線探索部8を備えている。また、縦直線抽出部10は、縦隣接投影部11、縦線分検出部12、縦線分統合部13、縦直線検出部14及び縦直線探索部15を備えている。

【0073】また、入れ子構造処理部19は、矩形内の横枠決定部20及び縦枠決定部21を備えている。連結パターン抽出部1は、複数の枠が配置される位置の相対的な関係に依存することなく、各パターンを安定に抽出するために、縦、横及び斜めの8方向のいずれかで繋がっている8連結の部分パターンを、ラベリングにより連結パターンとして抽出する。このラベリングにより得られる部分パターンは、文字が接触していない枠、枠に接触していない文字又は文字の一部、または枠に接触している文字のいずれかである。そして、これらの部分パターンを判別して枠を抽出する。また、ラベリングで得られた部分パターンのサイズが後に必要になるので、部分パターンを矩形近似して得られる矩形の角の座標をラベリングの処理中に算出しておく。

【0074】図2は、連結パターン抽出部1の動作の一

例を示す図である。連結パターン抽出部1は、図2

(a)に示す画像の入力パターンが入力されると、図2(b)に示すように該画像の中で一番大きな外接矩形をもつラベル24を抽出する。

【0075】マスク処理部2は、連結パターン抽出部1で抽出された連結パターンが、ある一定の大きさ以上である場合、画像から極端な斜め成分を省き、枠だけに存在する長い直線の抽出を容易にするための処理を行う。

【0076】すなわち、画像全体に対して横長、縦長の2種類のマスクによる走査を行う。そして、マスク内においてパターンが占める割合を算出し、その割合がある所定値より大きければそのマスク内を全てパターンとみなし、また該所定値以下であれば、マスク内のパターンを削除することにより、縦横成分を抽出する。

【0077】そして、複数の行または列が続いて、前記の割合が所定値より大きくなった場合、それらをまとめて矩形範囲をつくり、その中心線を処理結果とする。また、該マスク処理の結果得られる線分どうしの隙間が開かないようにするために、マスク処理を行う範囲が互いに重なりを持つように設定する。なお、原画像はマスク処理画像とは別に記憶しておく。

【0078】例えば、縦×横が1×6画素の矩形範囲を指定するマスクにより横成分を抽出する場合、図3に示すようにして原画像の処理が行われる。図3(a)は、互いに重なり合ったマスク「1」～「9」を示し、図3(b)の原画像に対してこれらのマスクを用いてマスク処理を行うことにより、図3(c)において黒丸で示された横成分の画素が抽出される。

【0079】図4は、マスク処理部2の動作を示すフローチャートである。同図において、まず、ステップS1に示すように、連結パターン抽出部1で得られた部分パターンの原画像全体に対して、横長及び縦長の2種類のマスクで走査を行う。

【0080】次に、ステップS2に示すように、マスク内のパターンの占める割合、すなわち、マスクの面積に対するマスク内のパターンの面積を算出し、この割合が、所定値より大きいかどうかの判断を行う。そして、マスク内のパターンの占める割合が所定値より小さいと判断された場合、ステップS3に進み縦成分又は横成分がないとして処理を終了する。

【0081】一方、ステップS2でマスク内のパターンの占める割合が所定値より大きいと判断された場合、ステップS4に進み、マスク内を全てパターンとみなし、該パターンを、全て、縦成分又は横成分とする。

【0082】次に、ステップS5に示すように、ステップS4で得られた成分が、上下の成分又は左右の成分と接しているかどうかの判断を行う。そして、上下の成分又は左右の成分と接していないと判断された場合、ステップS6に進み、ステップS4で得られた成分を処理結果として出力する。

【0083】一方、ステップS5で上下の成分又は左右の成分と接していると判断された場合、ステップS7に進み、互いに接している成分を統合する。そして、ステップS8において、ステップS7で統合した成分から矩形範囲を作成し、ステップS9において、ステップS8得られた矩形範囲の中心線を処理結果として出力する。

【0084】以上のマスク処理により、図5(a)に示す連結パターン24から、図5(b)に示すような斜め成分が除去され線幅が均一化されたパターン25が得られる。

【0085】横隣接投影部4及び縦隣接投影部11は、マスク処理されたパターンを横方向及び縦方向に複数に分割し、横方向及び縦方向に分割したそれぞれの範囲内*

$$P(i) = p(i-j) + \dots + p(i) + \dots + p(i+j) \quad (1)$$

なお、図6に示す例は、(1)式において $j=1$ とおいたものである。

【0087】図7は、横隣接投影部4及び縦隣接投影部11の動作を示すフローチャートである。図7において、まず、ステップS11で、マスク処理部2で得られた部分パターンを横方向及び縦方向に複数の部分に分割する。次に、ステップS12で、横方向及び縦方向のそれぞれの分割範囲内で投影値を算出する。

【0088】次に、ステップS13において、ステップS12で算出されたそれぞれの投影値に周囲の投影値を加算する。次に、ステップS14で、(1)式に基づいて、隣接投影値 $P(i)$ を算出する。

【0089】横線分検出部5及び縦線分検出部12は、部分パターンのマスク処理画像に対する隣接投影値に基づいて、横方向及び縦方向におけるある一定長さの線分又は直線の一部を、矩形近似により検出する。

【0090】すなわち、横隣接投影部4及び縦隣接投影部11で算出された隣接投影値と縦横それぞれの分割長との比が所定の閾値以上である部分を直線の候補が存在する位置とする。また、連続する複数の行又は列が所定の閾値以上となった場合は、それらの連続する複数の行又は列をまとめた矩形範囲を直線の候補が存在する位置とする。なお、この矩形近似により検出した一定長さの線分又は直線の一部を、以後「矩形線分」と呼ぶ。この線分検出処理により、例えば、図8(a)に示すような斜めに傾いた横線26から、図8(b)に示す横方向に分割された3つの矩形線分27が得られる。

【0091】ここで、隣接投影法ではなく、通常の投影法を用いた場合、図8(a)に示すような斜めに傾いた直線を検出することは不可能である。このため、斜めに傾いた直線を通常の投影法で検出するには、部分パターンの分割数を増やし分割長を短くする必要がある。しかし、部分パターンの分割長が短くなると、文字を構成している短い直線も多数検出され、文字と枠とを区別することが困難になる。これに対して、隣接投影法を用いることにより、比較的長い直線を正確に検出することがで

*でパターンの隣接投影を算出し、ある一定の長さの線分又は直線の一部を矩形近似により検出する。ここで、隣接投影とは、注目行又は注目列の投影値に周囲の行又は列の投影値を足し合わせたものである。このことにより、直線が傾いていて複数の行又は複数の列にわたっている場合においても、直線を正確に検出することができる。従って、隣接投影を用いることにより、ブロック枠などの大きな枠を検出する場合、その枠が傾いていても、枠を構成する直線を検出することが可能である。

10 【0086】例えば、図6において、 i 行の投影値を $p(i)$ とすると、隣接投影値 $P(i)$ は、(1)式により算出することができる。

きる。

【0092】図9は、横線分検出部5及び縦線分検出部12の動作を示すフローチャートである。図9において、まず、ステップS21で、横隣接投影部4及び縦隣接投影部11で算出された隣接投影値と縦横それぞれの分割長との比が所定のしきい値以上であるかどうかを判定する。そして、隣接投影値と縦横それぞれの分割長との比が所定のしきい値以上でないと判断された場合、ステップS22に進み、線分が存在しないものとみなす。

【0093】一方、ステップS21で隣接投影値と縦横それぞれの分割長との比が所定のしきい値以上であると判断された場合、ステップS23に進み、線分が存在するものとみなす。

【0094】次に、ステップS24において、ステップS23で線分とみなされたパターンが、その上下に存在する線分と接しているかどうかを判断する。そして、上記パターンが上下に存在する線分と接していないと判断された場合、ステップS25に進み、そのパターンを矩形線分とする。

【0095】一方、ステップS24において、ステップS23で線分とみなされたパターンがその上下に存在する線分と接していると判断された場合、ステップS26に進み、上記パターンとその上下に存在する線分とを統合する。そして、ステップS27で、ステップS26で統合した線分を矩形線分として検出する。

40 【0096】横線分統合部6及び縦線分統合部13は、横線分検出部5及び縦線分検出部12で検出された矩形線分のうち、近隣の矩形線分同士を統合して長い直線とするものである。例えば、図10(b)に示すように、途中で途切れていない矩形線分 x 、 y 、 z が接触しているか又は繋がっている場合、これらの矩形線分 x 、 y 、 z を統合して長い直線とする。また、図10(a)に示すように、矩形線分 x 、 y が繋がっていない場合、矩形線分 x 、 y の垂直方向の距離が隣接投影法において加算する行又は列数 j 以内ならば、矩形線分 x 、 y 、 z を統合して長い直線とする。

【0097】図11は、横線分統合部6及び縦線分統合部13の動作を示すフローチャートである。同図において、まず、ステップS31で、横線分検出部5及び縦線分検出部12で検出された矩形線分に対し、矩形線分同士の距離を算出する。次に、ステップS32で、ステップS31で算出された矩形線分同士の距離が、隣接投影法において加算する行又は列数*j*以内であるかどうかを判定する。そして、矩形線分同士の距離が隣接投影法において加算する行又は列数*j*以内でない場合、ステップS33に進み、矩形線分の統合を行わない。

【0098】一方、ステップS32において、矩形線分同士の距離が隣接投影法において加算する行又は列数*j*以内であると判断された場合、ステップS33に進み、矩形線分の統合を行う。

【0099】横直線検出部7及び縦直線検出部14は、横線分統合部6及び縦線分統合部13において得られた直線を矩形近似し、横枠又は縦枠を構成する直線の候補を検出する。例えば、図12(a)に示す統合された直線を矩形近似し、図12(b)で破線により示された直線28を得る。

【0100】横直線探索部8及び縦直線探索部15は、横直線検出部7及び縦直線検出部14で矩形近似された横線または縦線の直線について、横線の場合は左右端、縦線の場合は上下端を正確に検出するために、上記直線を構成するパターンの探索を行う。例えば、図13に示すように、横線である矩形近似された直線31に対して、該直線31を構成するパターン32の検索を行う。そして、以後の処理において、この検索で得られたパターン32の左右端33、34を、上記矩形近似された直線31の左右端35、36の代わりに用いる。

【0101】図14は、横直線探索部8の動作を示すフローチャートである。図14において、まず、ステップS41に示すように、横直線検出部7で近似した矩形範囲内のパターンの中で、最も細い部分のX座標を算出する。

【0102】次に、ステップS42で、ステップS41で算出したX座標におけるパターンの中心点を算出する。そして、ステップS43に示すように、ステップS42で算出したパターンの中心点を探索の開始点とする。ここで、探索の開始点をパターンの最も細い部分とするのは、最も細い部分は文字である可能性が低いからである。

【0103】次に、ステップS44で直線の探索方向を右に設定する。そして、ステップS45で、ステップS43で求めた開始点をパターン中の現在地と設定する。次に、ステップS46で、ステップS45で設定した探索の現在地が上記矩形範囲の内部であるかどうかの判定を行い、探索の現在地が矩形範囲の内部でない場合、ステップS51に進む。

【0104】一方、ステップS46で探索の現在地が矩形範囲の内部であると判定された場合、ステップS47に進み、探索の現在地からみて探索方向隣にパターンがあるかどうか判定する。ここで、探索の現在地からみて探索方向隣にパターンがあるとは、図15に示すように、パターン41からみて右方向隣の位置にパターン42が存在していることを意味している。そして、探索の現在地からみて探索方向隣にパターン42があると判定された場合、ステップS48に進み、探索方向隣にあるパターン42を探索の現在地とする。

【0105】一方、ステップS47で探索の現在地からみて探索方向隣にパターンがないと判定された場合、ステップS49に進み、探索の現在地からみて探索方向斜め隣にパターンがあるかどうか判定する。ここで、探索の現在地からみて探索方向斜め隣にパターンがあるとは、図15に示すように、パターン43からみて右方向斜め隣の位置にパターン44a又はパターン44bが存在していることを意味している。そして、探索の現在地からみて探索方向斜め隣にパターン44a、44bがあると判定された場合、ステップS50に進み、探索方向斜め隣にあるパターン44a、44bを探索の現在地とする。なお、探索方向斜め隣にあるパターン44a、44bが2つある場合はパターン44a、44bのどちらか一方を探索の現在地とする。

【0106】一方、ステップS49で探索の現在地からみて探索方向斜め隣にパターン44a、44bがないと判定された場合、ステップS51に進み、探索方向は右に設定されているかどうかを判定する。そして、探索方向は右に設定されていない場合、処理を終了する。

【0107】ステップS51で探索方向は右に設定されている場合、ステップS52に進み、探索方向を左に設定する。そして、探索方向を右に設定して行った処理と同様に、ステップS45～ステップS51の処理を繰り返す。ここで、探索方向を右に設定して行った処理を行う場合、探索の現在地からみて探索方向隣にパターンがあるとは、図15に示すように、パターン46からみて左方向隣の位置にパターン46が存在していることを意味している。また、探索の現在地からみて探索方向斜め隣にパターンがあるとは、図15に示すように、パターン47からみて左方向斜め隣の位置にパターン48a又はパターン48bが存在していることを意味している。

【0108】なお、縦直線探索部15の動作は、図14のフローチャートの処理において、横直線検出部7で検出されたパターンではなく縦直線検出部14で検出されたパターンを対象とし、また、探索方向を左右から上下に変更し、さらに、X座標をY座標に変更する以外は横直線探索部8の動作と同様である。

【0109】横点線検出部9及び縦点線検出部16は、小さいパターンがある一定の間隔で並んでいる時、点線が存在するとし、この点線に対して上述した直線と同様

に矩形近似を行うものである。点線を検出した後は実線、点線ともに、同じ直線として扱う。なお、横点線検出部9及び縦点線検出部16で対象とする点線は、規則的に点が並んでいる点線であり、一点鎖線などのように大きさの異なるパターンによって構成される点線は対象としていない。

【0110】以下、図16に示す点線を含む表を例にとり、横点線検出部9及び縦点線検出部16の動作を説明する。まず、図17に示すように、図16のパターンからラベリングにより算出されたラベルの外接矩形50内の画像の一部を、縦または横方向に短冊51として取り出す。そして、その短冊51の内部に存在するパターンの大きさを、予め設定していた点線を構成するパターンの大きさと比較し、上記予め設定していた点線を構成するパターンと同じ大きさを有する部分パターンを探す（ここでは例として縦方向の点線を抽出する場合を示す）。

【0111】次に、図18に示すように、予め設定してあった大きさと同じ大きさの部分パターンが見つかった部分を、図17の短冊51に垂直な（この場合は縦方向に）短冊53として切り出す。そして、この短冊53の範囲内で、上記予め設定していた点線を構成するパターンと同じ大きさを有する部分パターンを抽出する。

【0112】次に、図19に示すように、上記抽出された部分パターンが、ある一定の間隔で並び、かつある一定の数以上連続して並んでいるものを点線部と判断して、これを矩形54により近似する。この場合、点線部の両端の点を実線と重なっている場合も考慮して、該矩形54を設定する際、その上下端を、最上位の点と最下位の点の位置から点と点の間隔分の距離だけ伸ばすようにする。

【0113】そして、図20に示すように、以上の処理を繰り返して、次々に点線を抽出し、これらを矩形55により近似する。以下、抽出済の実線と、新たに抽出された点線は同様なものとして扱う。よって、以後、縦線は縦方向の実線と点線とを示し、横線は横方向の実線と点線とを示す。

【0114】図21は、横点線検出部9及び縦点線検出部16の動作を示すフローチャートである。図21において、まず、ステップS51で、入力された画像の一部を横方向又は縦方向に短冊状に切り出す。

【0115】次に、ステップS52で、ステップS51で短冊状に切り出した範囲内で、ある一定の大きさの部分パターンが存在するかどうかを判定する。そして、ある一定の大きさの部分パターンが存在しない場合、処理を終了する。

【0116】一方、ステップS52において、ある一定の大きさの部分パターンが存在すると判定された場合、ステップS53に進み、抽出した部分パターンを含み、且つステップS51で用いた短冊と垂直な短冊状の画像

を入力画像から切り出す。

【0117】次に、ステップS54で、ステップS53で切り出した短冊状の画像の内部に、ある一定の大きさの部分パターンが存在するかどうかを判定する。そして、ある一定の大きさの部分パターンが存在しない場合、処理を終了する。

【0118】一方、ステップS54において、ある一定の大きさの部分パターンが存在すると判定された場合、ステップS55に進み、上記部分パターンを抽出し、該部分パターンが一定の間隔で並んでいるかどうか判定する。そして、該部分パターンが一定の間隔で並んでいない場合、処理を終了する。

【0119】一方、ステップS55において、上記抽出した部分パターンが一定の間隔で並んでいると判定された場合、ステップS56に進み、部分パターンの存在する範囲を矩形で囲む。

【0120】次に、ステップS57で、点と点との間隔分の長さだけステップS56で得られた矩形の両端を延ばす。そして、ステップS58で、該矩形で囲まれた範囲を点線の存在範囲とする。

【0121】図22に、上述した横直線抽出部3で抽出された横線56と縦点線検出部16で検出された縦線57の例を示す。次に、以上の処理により抽出された直線に基づいて、枠を検出する処理を行う。この枠検出処理は、上述の処理により抽出された縦横の直線の中から、枠を構成する直線を決定し、上下左右の四辺を枠で囲まれた矩形範囲を左上から順番に抽出することで、枠の構造を検出するものである。

【0122】例えば、図23(a)に示す表60の場合、まず、該表60を、図23(b)に示すように、大枠の複数の行に分割する。次に、図23(c)に示すように、該行をさらに小さい矩形に分割する。尚、この段階で検出された矩形がさらに細い矩形に分割されている状態を「入れ子」と呼び、この「入れ子」状態にある矩形については新たな表とみなして、図23(d)に示すように該矩形をさらに小さな矩形に分割する。

【0123】横枠決定部17は、上記検出された横線の中から、各行の横枠を構成している2本の横線の組を決定する処理を実行するものである。この処理は以下のような規則に従って行われる。

【0124】1) 横直線抽出部3で検出された横直線（横線）の中から、上から順番に横線を2本抽出し、これらを2本の横線の組の候補とする。

2) 上記抽出した2本の横線が、長さが同じ、又は下の直線の方が長ければ、その2本を、横線の組とする。このとき、2本の横線の長さが異なる場合、長い方の横線を再使用可とする。

【0125】3) 上記抽出した2本の横線の内、下の横線の方が短かければ、さらにその下の横線を横枠の候補とし、これらの横線の長さを比較する。また、下の横線

の方が右方向又は左方向に短かければ、さらにその下の直線を組の候補とし、これらの横線の長さを比較する。

【0126】4) 該比較の結果、上記条件2) を満足する直線が下部に無いことが分かれば、この場合に限り、上記上の横線とその下の横線とを組みとする。

5) 一番下の横線の処理後、上部にまだ未処理の直線がある場合は、再び上から順に未処理の横線と再使用可の横線とを用いて、上記1)～4) の処理を行い、2本の横線の組を作る。

【0127】以上のようにして決定される組となった2本の横線は、以後の処理において、行などの矩形枠を構成する横枠の組(2本組の横枠)として取り扱われる。図24は、横枠決定部17における上記横線の組みを決定する処理を説明する示すフローチャートである。

【0128】図24において、まず、ステップS61で、横直線抽出部3で検出された横直線の中から、最上部の2本を取り出す。次に、ステップS62で、上記取り出した2本の横直線のうち、上側の横直線をupline、下側の横直線をbllineとする。

【0129】次に、ステップS63で、直線uplineと直線bllineとに縦方向の重なりがあるかどうかを判定する。そして、直線uplineと直線bllineとに縦方向の重なりがない場合、ステップS64に進み、直線bllineの下の直線を新たにbllineとし、この処理を直線uplineと直線bllineとに縦方向の重なりがあるようになるまで続ける。

【0130】一方、ステップS63において、直線uplineと直線bllineとに縦方向の重なりがある場合、ステップS65に進み、直線uplineと直線bllineとの長さが同じであるか又は直線uplineより直線bllineの方が長いかどうかを判定する。そして、上記条件を満足する直線uplineと直線bllineが存在する場合、ステップS68に進み、上記直線uplineと上記直線bllineを2本の横線の組(2本組の横枠)とする。

【0131】一方、ステップS65において、直線uplineと直線bllineとの長さが異なり、且つ、直線uplineより直線bllineの方が短い場合、ステップS66に進み、直線bllineより下方に直線uplineと長さが同じであるか又は直線uplineより長い直線blline2が存在するかどうかを判定する。そして、上記条件を満足する直線bllineと直線uplineが存在する場合、ステップS67に進み、直線bllineを直線blline2に変更してステップS68に進む。

【0132】ここで、直線uplineと直線bllineとの長さが同じであるとは、図26(a)のような状態を示し、直線uplineより直線bllineの方が短いとは、図26(b)のような状態を示し、直線uplineより直線bllineの方が長いとは、図

26(c)のような状態を示す。

【0133】一方、ステップS66において、直線bllineより下方に直線uplineと長さが同じであるか又は直線uplineより長い直線blline2が存在しない場合、ステップS68に進み、直線uplineと直線bllineとを2本組の横枠とする。

【0134】次に、ステップS69で、後述する縦枠決定処理と入れ子処理を行う。次に、ステップS70で、ステップS68で2本組の横枠とした直線uplineと直線bllineの長さが異なるかどうかを判定する。そして、直線uplineと直線bllineの長さが異なる場合、ステップS71に進み、右方向へ長い直線又は左方向に長い直線を再利用可として、ステップS72に進む。

【0135】ここで、上記右方向へ長い直線又は左方向に長い直線とは、図26(d)に示す直線blline、同図(e)に示す直線uplineと直線bllineである。

【0136】一方、ステップS70において、直線uplineと直線bllineとの長さが同じであると判定された場合、ステップS72に進み、直線bllineより下方に直線が存在するかどうかを判定する。そして、直線bllineより下方に直線が存在する場合、ステップS73に進み、直線bllineとその下の直線を抽出し、ステップS62に戻る。

【0137】一方、ステップS72において、直線bllineより下方に直線が存在しないと判定された場合、ステップS74に進み、再使用可の直線を除いて、処理済の直線を処理対象からはずす。

【0138】次に、ステップS75で、未処理の直線が存在するかどうかを判定する。そして、未処理の直線が存在しない場合、処理を終了する。一方、ステップS75において、未処理の直線が存在すると判定された場合、ステップS76に進み、最上部から直線を2本抽出し、ステップS62に戻る。

【0139】次に、上記横枠の2本組を決定する動作を、図27に示す帳票70を例にとって説明する。上記帳票70では、横直線①と横直線③、横直線③と横直線⑤、横直線⑤と横直線⑥、横直線⑥と横直線⑦、横直線⑦と横直線⑧、横直線②と横直線④、横直線④と横直線⑤が、それぞれ横枠の2本組となる。ここで、上記①～⑧の横直線(横線)は、横直線抽出部3で抽出される。また、I～Vの縦直線(縦線)は縦直線抽出部10で抽出される。

【0140】この様な形式の帳票70から上記横枠の2本組を抽出する場合、まず、最上部の2本の横直線①、③を取り出す。そして、この2本の横直線①、③は同じ長さであるので組にする。

【0141】次に、横直線③の下の横直線を探索し、2本の横直線③、⑤を取り出す。そして、下側の横直線⑤

は上側の横直線③より長いので、横直線③と横直線⑤とを組にする。また、このとき、長い方の横直線⑤を再使用可とする。

【0142】次に、横直線⑤の下横直線を探索し、2本の横直線⑤、⑥を取り出す。そして、この2本の横直線⑤、⑥は同じ長さであるので、横直線⑤と横直線⑥とを組にする。

【0143】次に、横直線⑥の下横直線を探索し、2本の横直線⑥、⑦を取り出す。そして、この2本の横直線⑥、⑦は同じ長さであるので、横直線⑥と横直線⑦とを組にする。

【0144】次に、横直線⑦の下横直線を探索し、2本の横直線⑦、⑧を取り出す。そして、この2本の横直線⑦、⑧は同じ長さであるので、横直線⑦と横直線⑧とを組にする。

【0145】次に、横直線②、④が未使用なので一番上に戻り、2本の横直線②、④を取り出す。そして、この2本の横直線②、④は同じ長さであるので組にする。次に、横直線④の下横直線を探索し、横直線④と再使用可に設定されている横直線⑤とを取り出す。そして、下側の横直線⑤は上側の横直線④より長いので、横直線④と横直線⑤とを組にする。

【0146】次に、図28に示す帳票80において、2本組の横枠を決定する動作を説明する。この帳票80の場合、横直線①と横直線②、横直線②と横直線③、横直線③と横直線④、横直線④と横直線⑤がそれぞれ組となる。尚、上記横直線①～⑤は横直線抽出部3で抽出される。また、縦直線Ⅰ～Ⅳは縦直線抽出部10で抽出される。

【0147】上記帳票80の場合、まず、最上部の2本の横直線①、②を取り出す。そして、この2本の横直線①、②は同じ長さであるので組にする。次に、横直線②の下横直線を探索し、横直線③を取り出す。ここで、横直線②と横直線③とは長さが異なり、下側の横直線③は上側の横直線②より長さが短いので、横直線③の下横直線④、⑤の長さを調べる。その結果、横直線③の下には横直線②と同じ長さの横直線は存在しないため、初めに抽出した横直線②と横直線③とを組にする。

【0148】次に、横直線③と横直線④とを取り出す。ここで、横直線③と横直線④とは長さが異なり、下側の横直線④は上側の横直線③より長さが短いので、横直線④の下横直線⑤の長さを調べる。その結果、横直線④の下には横直線③と同じ長さの横直線は存在しないため、初めに抽出した横直線③と横直線④とを組にする。

【0149】次に、横直線④と横直線⑤とを取り出す。ここで、横直線④と横直線⑤とは長さが異なり、下側の横直線⑤は上側の横直線④より長さが短いので、横直線⑤の下横直線の長さを調べる。その結果、横直線⑤の下には横直線④と同じ長さの横直線は存在しないため、初めに抽出した横直線④と横直線⑤とを組にする。

【0150】以上で、未処理の横直線が存在しなくなったため、処理を終了する。次に、図29に示す帳票90において、2本組の横枠を決定する動作を説明する。

【0151】上記帳票90の場合、横直線⑤と横直線⑥、横直線⑥と横直線②、横直線②と横直線⑦、横直線⑦と横直線⑧、横直線①と横直線②、横直線②と横直線③、横直線③と横直線④が、それぞれ、2本組の横枠となる。ここで、横直線①～⑧は横直線抽出部3で抽出される。また、縦直線Ⅰ～Ⅳは縦直線抽出部10で抽出される。

【0152】この場合、横直線①～⑧を上から検索し、まず、最上部の2本の横直線⑤、①を取り出す。この2本の横直線⑤、①は、縦方向に重なり部分がないため、次に、横直線①の下横直線⑥を取り出す。そして、この2本の横直線⑤、⑥は縦方向に重なり部分を有し、且つ、同じ長さであるので組にする。

【0153】次に、横直線⑥の下横直線を探索し、2本の横直線⑥、②を取り出す。そして、下側の横直線②は上側の横直線⑥より長いので、横直線⑥と横直線②とを組にする。この時、長い方の横直線②を再使用可として記憶する。

【0154】次に、横直線②の下横直線を探索し、2本の横直線②、⑦を取り出す。ここで、横直線②と横直線⑦とは長さが異なり、下側の横直線⑦は上側の横直線②より長さが短いので、横直線⑦より下の横直線③、④、⑧の長さを調べる。その結果、横直線②と同じ長さの横直線は横直線⑦の下に存在しないため、初めに抽出された横直線②と横直線⑦とを組にする。

【0155】次に、横直線⑦の下横直線を探索し、2本の横直線⑦、③を取り出す。しかし、この2本の横直線⑦、③は、縦方向に重なり部分がないため、横直線③の下横直線⑧を取り出す。そして、この2本の横直線⑦、⑧は同じ長さであるので組にする。

【0156】次に、横直線⑧の下横直線を探索し、2本の横直線⑧、④を取り出す。しかし、この2本の横直線④、⑧は、縦方向に重なり部分がないため、組を作ることが不可能である。以上により、処理が一番下の直線まで到達したため、再び未処理の直線又は再利用可の直線を上部から2本抽出する。

【0157】すなわち、未処理の横直線①及び再利用可の横直線②を取り出す。そして、下側の横直線②は上側の横直線①より長いので、横直線①と横直線②とを組にする。

【0158】次に、横直線②の下横直線を探索し、2本の横直線②、③を取り出す。ここで、横直線②と横直線③とは長さが異なり、下側の横直線③は上側の横直線②より長さが短いので、横直線③の下横直線④、⑧の長さを調べる。その結果、横直線②と同じ長さの横直線は横直線③の下に存在しないため、初めに抽出された横直線②と横直線③とを組にする。

【0159】次に、横直線③の下横直線を探索し、2本の横直線③、④を取り出す。そして、横直線④は横直線③と同じ長さであるので、横直線③と横直線④とを組にする。以上により未処理の横直線が存在しなくなったため、処理を終了する。

【0160】次に、図30に示す形式の帳票100から横枠の2本組を抽出する処理を説明する。上記帳票100の場合、横直線①と横直線②、横直線②と横直線③、横直線③と横直線④、横直線⑤と横直線⑥、横直線⑥と横直線③がそれぞれ組となる。ここで、横直線①～⑥は横直線抽出部3で抽出される。また、縦直線Ⅰ～Ⅴは縦直線抽出部10で抽出される。

【0161】この場合、まず、横直線①～⑥を上から検索し、最上部の2本の横直線①、⑤を取り出す。しかし、この2本の横直線①、⑤は、縦方向に重なり部分がないため、その下の横直線②を取り出す。そして、この2本の横直線①、②は同じ長さであるので組にする。

【0162】次に、横直線②の下横直線を探索し、2本の横直線②、③を取り出す。ここで、横直線②と横直線③とは長さが異なり、図26(d)に示すように下側の横直線③が上側の横直線②より左方向に長さが短いので、横直線③の下横直線④の長さを調べる。その結果、横直線②と同じ長さの横直線は横直線③の下に存在しないため、初めに抽出された横直線②と横直線③とを組にする。このとき、横直線②と横直線③は、“右方向又は左方向に長さが短い関係にある”ので、横直線②と横直線③とを再使用可として記憶する。尚、“右方向又は左方向に長さが短い”状態にある直線とは、図26(d)、(e)に示すような配置関係にある2本の直線のことである。

【0163】次に、横直線③の下横直線を探索し、2本の横直線③、④を取り出す。そして、横直線③と横直線④とは同じ長さであるので組にする。以上により、処理が一番下の直線まで到達したため、再び未処理の直線又は再利用可の直線を上部から2本抽出する。

【0164】すなわち、未処理の2本の横直線⑤、⑥を上から取り出す。そして、横直線⑤と横直線⑥とは同じ長さであるので組にする。次に、横直線⑥の下横直線を探索し、2本の横直線⑥、③を取り出す。そして、そして、下側の横直線③は上側の横直線⑥より長いので、横直線③と横直線⑥とを組にする。以上により未処理の横直線が存在しなくなったため、処理を終了する。

【0165】縦枠決定部18は、枠の検出処理において縦枠を決定する処理を行うもので、縦直線抽出部10で抽出された縦線のうち、横枠決定部17で横枠の2本組として決定された2本の横線に上下とも達しているものを縦枠と決定する。

【0166】矩形表現部22は、横枠決定部17で決定された2本組の横枠と縦枠決定部18で決定された該2本組の横枠の上下に達している2本の縦枠とで囲まれた

矩形を抽出するものである。

【0167】入れ子構造処理部19は、矩形表現部22で抽出された矩形の中から、その内部にさらに矩形を含む入れ子構造の矩形を検出し、この矩形をさらに小さな矩形に分割するこの入れ子構造の矩形の検出は、矩形表現部22で抽出された矩形について、その内部に、該矩形の左右の2辺に達している横線が存在するか否かを調べることにより行う。例えば、図31(a)に示すように、抽出された矩形100が、その内部に左右の2辺101、102に達している横線が存在する場合、この矩形を入れ子構造の矩形とみなす。一方、図31(b)に示す矩形120のように、その内部に横線123を含んでいても、該横線123がの両端又はいずれか一端が、該矩形120左右の2辺121または122のいずれかに達していない場合、この横線は枠の一部である可能性が低いため、入れ子構造の矩形とはみなさない。

【0168】図32は、上述した図24のステップS69の処理に該当する縦枠決定部18による縦枠決定処理及び入れ子構造処理部19による入れ子処理を説明するフローチャートである。

【0169】同図において、まず、ステップS80で最上位の行の2本組の横枠を決定(選択)し、次に、ステップS81で、縦直線抽出部10で抽出した縦直線のうち、上端と下端との両方が上記2本組の横枠に達しているものを選択する。

【0170】次に、ステップS82で、上記選択した縦直線を縦枠と決定する。次に、ステップS83で、上記2本組の横枠と上記2本の縦枠とで囲まれた矩形範囲(矩形)を抽出する。

【0171】次に、ステップS84で、上記矩形の左右2本の縦枠に両端が達している横直線が存在するかどうか判定する。そして、そのような横直線が存在しない場合、次の行の横枠を決定(選択)する処理を行い、これにより新たに決定された2本組の横枠に対して上記ステップS81～S86と同様の処理を実行する。

【0172】一方、ステップS84において、左右2本の縦枠に両端の達している横直線が存在すると判定された場合、ステップS85に進み、上記矩形を新たに小さな表(入れ子構造の矩形)とみなし、上述した図24のフローチャートに示す方法により、該入れ子構造の矩形内に存在する小さな矩形の横枠を決定する。この横枠の決定処理は横枠決定部20により行われる。

【0173】次に、ステップS86で、縦枠決定部21により、上記矩形の縦枠を決定する処理を行う。該縦枠決定部21は、上記縦枠決定部18と同様な方法により縦枠を決定する。

【0174】上記動作は、横枠決定部17により決定された全ての行の2本組の横枠に対して実行され、帳票の全ての矩形について、横枠と縦枠が決定される。次に、図1にシステム構成が示された本実施例の全体的な動作

の流れを、図33～図36を参照しながら説明する。

【0175】以下の説明では、図33に示すパターンを有する帳票110から矩形枠を全て抽出する本実施例の動作について説明する。尚、上述した横直線抽出部3及び縦直線抽出部10の動作によって、上記帳票110を構成する全ての横線と縦線が抽出されいるとの前提の元で説明を行う。

【0176】図33～図36において、①～⑥は横直線（横線）を示し、Ⅰ～Ⅴは縦直線（縦線）を示している。

〔処理1〕まず、横枠決定部17によって横枠決定処理を行う。これは、上述した横線の組を決定する方法に従って、横直線抽出部3で検出された全ての横線について、矩形枠を構成する2本の横線の組を決定していく処理である。このとき、組となった横線を横枠とし、一番上の行から順番に1行づつ処理していく。これにより、図33に示すように、まず、横直線①と横直線②とを2本組の横線（組の横線）と判断し、該横直線①と横直線②を横枠とする。

【0177】〔処理2〕次に、縦枠決定部18によって縦枠決定処理を行う。これは、縦直線抽出部10で検出された縦線のうち、上記横枠決定処理によって組と判断された2本の横線に上下とも達しているものを縦枠とする処理である。これにより、例えば、図34に示すように、縦直線Ⅰと縦直線Ⅴが組となっている横線に上下とも達しているので、縦直線Ⅰと縦直線Ⅴとを縦枠とする。

【0178】〔処理3〕次に、矩形表現部22によって矩形表現処理を行う。これは、横枠決定部17により決定された組となっている横枠2本と縦枠決定部18により決定された縦枠2本とで囲まれた矩形範囲を、矩形枠として抽出するものである。例えば、図35に示すように、横直線①と縦直線ⅠⅤと横直線②と縦直線Ⅰとで囲まれた範囲及び横直線①と縦直線ⅤⅠと横直線②と縦直線ⅠⅤとで囲まれた範囲を矩形範囲とする。

【0179】〔処理4〕次に、入れ子構造処理部19によって入れ子処理を行う。これは、上記矩形表現部22によって抽出された矩形範囲内に横線が存在する場合、該矩形範囲を新たに表とみなし、この表に対して、横枠決定処理、縦枠決定処理及び矩形表現処理を行うものである。例えば、図36に示すように、横直線②と縦直線ⅠⅠⅠと横直線④と縦直線Ⅰとで囲まれた矩形範囲は入れ子状態となっており、その中に3個の矩形を含んでいる。このため、この入れ子状態の矩形範囲を、横直線②と縦直線ⅠⅠⅠと横直線③と縦直線Ⅰとで囲まれた矩形範囲、横直線③と縦直線ⅠⅠと横直線④と縦直線Ⅰとで囲まれた矩形範囲及び横直線③と縦直線ⅠⅠⅠと横直線④と縦直線ⅠⅠとで囲まれた矩形範囲の3領域に分割する。

【0180】以後、上述した〔処理1〕～〔処理4〕

を、残りの横線と縦線について、次の行から繰り返す。これにより、図33に示す帳票110は、それが有する全ての矩形領域に分割される。

【0181】ところで、図37(a)に示すように矩形で表現される点線領域を、図37(b)に示すように、さらに細かく分割して小矩形で表現するようにしてもよい。このとき、小矩形の長さは、例えば、上述した隣接投影法で抽出する線分の長さと同しくなるように設定する。

10 【0182】また、既に、決定された縦枠と横枠により矩形領域を表現した後、上述した隣接投影法により抽出される線分の位置情報を用いて、該矩形領域内の枠領域の位置を詳細に表現するようにしてもよい。例えば、図38(a)において破線121で示す矩形領域の位置情報と、図38(b)において実線122で示す上記隣接投影法により得られる矩形領域に統合する前の直線の位置情報とを用いて、図38(c)において破線123で示すように上記矩形領域内の枠領域の位置を詳細に表現するようにしてもよい。

20 【0183】以上、説明したように、本実施例によれば、帳票を枠で囲まれた領域に分割することができるので、様々な形式の帳票を処理することが可能となる。ただし、本実施例で対象となる枠領域は、矩形形状のものであり、図39に示すように矩形範囲で表現できない形状のものは対象外である。

【0184】図40と図41に、本実施例の実験例を示す。図40は、この実験でサンプルとして用いた帳票140の原画像を示す図である。

30 【0185】また、図41は、上記帳票140を下記に示す入力パターン条件の基で処理した結果を示す図である。

入力パターンの条件

解像度 : 100dpi
 大きさ : 横 848ドット
 : 縦 1169ドット
 かすれ : なし
 画素 : 2値
 罫線の種類 : 実線、点線

40 このように、本実施例によれば、不規則な構造を持ち、実線と点線が混在している帳票であっても、該帳票から矩形枠を正確に抽出することが可能であることが確かめられた。

【0186】以上説明したように、本発明の一実施例によるパターン抽出装置によれば、様々な構造の表形式枠から直線及び点線を抽出し、この抽出された直線及び点線が枠であるかを判断し、その判断の結果得られた枠情報を基に、文字が書かれる文字記入枠を正確に抽出することが可能となる。

50 【0187】また、横直線抽出部3及び縦直線抽出部10において、横隣接投影部4及び縦横隣接投影部11を

設け、ある行又は列の投影値に周囲の行又は列の投影値を足し合わせた値をその行又は列の投影値とする隣接投影法により線分を抽出するようにしたので、斜めに傾いた直線を正確に検出することができる。

【0188】また、横直線抽出部3及び縦直線抽出部10において、横直線探索部8及び縦直線探索部15を設け、矩形近似された直線の最も細い部分から外側へ探索を行って直線の両端の位置を検出することにより、直線の両端の位置を正確に検出することが可能となる。

【0189】また、横点線検出部9及び縦点線検出部16において、全画像の一部の領域の連結情報に基づき、部分パターンのサイズと部分パターン間の距離より点線を抽出することにより、点線を抽出する処理時間を短縮することが可能となる。

【0190】また、点線が存在する部分を矩形で表現することにより、点線を矩形で近似された実線と同様に扱うことが可能となる。また、矩形で表現された点線部分を、図37に示すように、さらに細かい矩形で表現することにより、点線が傾いていても、正確に該点線が存在する位置を表現することが可能となる。

【0191】また、矩形近似された点線を直線と同様な情報を用いて表現することにより、点線と実線を同様に扱うことが可能となる。また、横枠決定部17及び矩形内の横枠決定部20において、抽出された横方向の直線又は点線のうち、ある一定の長さより長いものを全て横枠の候補とすることで、枠の大きさと比較すると非常に短い横線であっても、候補として抽出することが可能となる。

【0192】また、横枠決定部17及び矩形内の横枠決定部20において抽出された直線又は点線を、上から順番に2本づつ組としながら、1行づつ横枠を決定する処理を続けていくことにより、外形が矩形でない枠についても枠の抽出処理に対応することが可能となる。

【0193】また、入れ子構造処理部19により、入れ子状態の矩形をさらに小さな矩形に分割するので、複雑な構造をもつ表であっても、枠の抽出や文字領域の抽出が可能となる。

【0194】また、横隣接投影部4及び縦横隣接投影部11において、対象画像の解像度に基づいて隣接投影の範囲を変化させることにより、直線をより正確に抽出することが可能となる。

【0195】また、横枠決定部17及び縦枠決定部18において、抽出された直線または点線が枠であるかを判断する場合、枠の候補の長さを比較する際に用いる閾値を線の種類により変化させることにより、実線と点線が混在している場合であっても正確に枠を決定することが可能となる。

【0196】また、横点線検出部9及び縦点線検出部16において、点線を矩形で表現する際に、該点線の外接矩形の両端を、点線を構成する点の1間隔分の長さだけ

伸ばすことにより、点線の位置を、より正確に記述することが可能となる。

【0197】また、矩形表現部22において、縦枠と横枠により文字領域を表現する方法に加えて、隣接投影法によって抽出された線分でより細かく文字領域の位置を表現することで、文字パターンの変形を防ぐことが可能となる。

【0198】

【発明の効果】以上説明したように、請求項1の発明によれば、枠を構成する直線情報に基づいて、矩形枠を抽出することにより、様々な構造の表形式枠を抽出することができる。

【0199】また、請求項2の発明によれば、矩形枠内に一定以上の長さを有する横方向の直線が存在する場合、その矩形領域を新たに表とみなし、その表内に存在する矩形枠を再度抽出することにより、複雑な入れ子構造を有する表であっても、枠の抽出や文字領域の抽出が可能となる。

【0200】また、請求項3の発明によれば、直線及び点線が枠を構成するかどうかを判断することにより、点線と実線とが混在している表の枠を正確に抽出できる。また、請求項4の発明によれば、直線及び点線から構成される矩形枠内に一定以上の長さを有する横方向の直線または点線が存在する場合、その矩形領域を新たに表とみなし、その表内に存在する矩形枠を抽出することにより、点線と実線とが混在している入れ子構造の表であっても、枠の抽出や文字領域の抽出が可能となる。

【0201】また、請求項5の発明によれば、周囲の行又は列の投影値を足し合わせた投影値により矩形近似して直線を検出することにより、斜めに傾いた直線を正確に検出することができる。

【0202】また、請求項6の発明によれば、直線の原画像のパターンの最も細い部分から外側へ探索することにより、直線の両端の位置を正確に検出することが可能となる。

【0203】また、請求項7の発明によれば、矩形枠の領域の位置情報を矩形枠に統合される前の線分の位置情報を用いて、より細かく設定することにより、直線の位置をより正確に検出することが可能となる。

【0204】また、請求項8の発明によれば、隣接投影の範囲を対象画像の解像度に基づいて変化させることにより、直線の抽出をより正確に行うことが可能となる。また、請求項9の発明によれば、部分パターンのサイズ及び部分パターン間の距離により点線を抽出することにより、処理時間の短縮を図ることが可能となる。

【0205】また、請求項10の発明によれば、点線が存在する領域を矩形に近似して表現することにより、点線を矩形で近似された実線と同様に扱うことが可能となる。また、請求項11の発明によれば、矩形により表現された点線領域をさらに細分割して小矩形で表現するこ

とにより、傾きがあっても正確に点線の存在する位置を表現することが可能となる。

【0206】また、請求項12の発明によれば、矩形により表現された点線を直線と同様な位置情報を用いて表現することにより、点線を実線と同様に扱うことが可能となる。

【0207】また、請求項13の発明によれば、横方向の直線の中から、ある一定の長さ以上の直線を抽出して横枠線の候補とすることにより、枠の大きさと比較すると非常に短い横線であっても、枠の候補として抽出することが可能である。

【0208】また、請求項14の発明によれば、枠の候補の直線の長さを比較する際に用いる閾値を、直線の種類により変化させることにより、実線と点線の混在している場合であっても正確に枠を決定することが可能となる。

【0209】また、請求項15の発明によれば、横枠線の候補の直線を、上部から順番に2本ずつ取り出して組とすることにより、長さが異なる横直線を横枠の組として処理することができ、枠の外形が矩形でない場合にも

枠の抽出が可能となる。

【0210】また、請求項16の発明によれば、2本の組の横方向の直線がほぼ同じ長さである場合、この2本の横方向の直線を、横枠の組として登録することにより、枠の外形が矩形の場合の枠の抽出が可能となる。

【0211】また、請求項17の発明によれば、2本の組の横方向の直線のうち、下側の直線の方が長い場合、この2本の横方向の直線を、横枠線の組として登録することにより、枠の外形が矩形でない場合にも枠の抽出が可能となる。

【0212】また、請求項18の発明によれば、2本の横方向の上側の第1の直線及び下側の第2の直線において、前記第1の直線より前記第2の直線の方が短く、且つ、前記第2の直線の下方に前記第1の直線と同じ長さ又は前記第1の直線より長い横方向の第3の直線が存在しない場合、前記第1の直線及び前記第2の直線を、横枠線の組として登録することにより、枠の外形が矩形でない場合にも枠の抽出が可能となる。

【0213】また、請求項19の発明によれば、2本の横方向の上側の第1の直線及び下側の第2の直線において、前記第1の直線より前記第2の直線の方が短く、且つ、前記第2の直線の下方に前記第1の直線と同じ長さ又は前記第1の直線より長い横方向の第3の直線が存在する場合、前記第1の直線及び前記第3の直線を、横枠線の組として登録することにより、枠の外形が矩形でない場合にも枠の抽出が可能となる。

【0214】また、請求項20の発明によれば、2本の横方向の上側の第1の直線及び下側の第2の直線において、前記第1の直線に対して前記第2の直線が右方向又は左方向に短く、且つ、前記第2の直線の下方に前記第

1の直線と同じ長さ又は前記第1の直線より長い横方向の第3の直線が存在しない場合、前記第1の直線及び前記第2の直線を、横枠線の組として登録することにより、枠の外形が矩形でなく、且つ、上側の直線と下側の直線とが横方向にずれて配置されている場合にも枠の抽出が可能となる。

【0215】また、請求項21の発明によれば、2本の横方向の上側の第1の直線及び下側の第2の直線において、前記第1の直線に対して前記第2の直線が右方向又は左方向に短く、且つ、前記第2の直線の下方に前記第1の直線と同じ長さ又は前記第1の直線より長い横方向の第3の直線が存在する場合、前記第1の直線及び前記第3の直線を、横枠線の組として登録することにより、枠の外形が矩形でなく、且つ、上側の直線と下側の直線とが横方向にずれて配置されている場合にも枠の抽出が可能となる。

【0216】また、請求項22の発明によれば、2本の横方向の上側の第1の直線及び下側の第2の直線において縦方向の重なりがない場合、前記第2の直線に代えて前記第2の直線の下方に存在する横方向の第3の直線を、横枠線の候補とすることにより、枠の外形が矩形でなく、且つ、上側の直線と下側の直線とが縦方向にずれて配置されている場合にも枠の抽出が可能となる。

【0217】また、請求項23の発明によれば、下側の横枠が右方向へ長い又は左方向へ長い場合、横枠線の両方を横枠線の候補として、再度枠判断処理のために選択可能とすることにより、枠の外形が矩形でなく、且つ、上側の直線と下側の直線とが横方向にずれて配置されている場合においても、未処理の横枠の組を作ることが可能となる。

【0218】また、請求項24の発明によれば、長い方の横枠を横枠の候補として再度枠判断処理のために選択可能とすることにより、未処理として残った横線を横枠の組とすることが可能となる。

【0219】また、請求項25の発明によれば、未処理の横枠の候補が存在している場合、再び部分パターンの上部から未処理の横枠を2本ずつ組として枠の候補として抽出することにより、未処理として残った横線を横枠の組とすることが可能となる。

【0220】また、請求項26の発明によれば、2本の横枠に両端が達している縦線を縦枠として登録することにより、縦枠を抽出する場合において、文字と枠とを正確に区別することができる。

【0221】また、請求項27の発明によれば、横方向の直線または点線の中から、ある一定の長さ以上の線分を抽出し、その線分を横枠線の候補とすることにより、横枠を抽出する場合において、文字と枠とを正確に区別することができる。

【0222】また、請求項28の発明によれば、横枠線の候補の直線または点線を、上部から順番に2本ずつ取

り出して組とすることにより、長さが異なる横直線または横点線を横枠の組として処理することができ、枠の外形が矩形でなく、且つ、点線が混在している場合にも枠の抽出が可能となる。

【0223】また、請求項29の発明によれば、2本の組の横方向の直線がほぼ同じ長さである場合、この2本の横方向の直線を、横枠線の組として登録することにより、枠の外形が矩形で点線が混在している場合の枠の抽出が可能となる。

【0224】また、請求項30の発明によれば、2本の組の横方向の直線または点線のうち、下側の直線または点線の方が長い場合、この2本の横方向の直線または点線を、横枠線の組として登録することにより、枠の外形が矩形でなく、且つ、点線が混在している場合にも枠の抽出が可能となる。

【0225】また、請求項31の発明によれば、2本の横方向の上側の第1の直線または点線及び下側の第2の直線または点線において、前記第1の直線または点線より前記第2の直線または点線の方が短く、且つ、前記第2の直線または点線の下方に前記第1の直線または点線と同じ長さ又は前記第1の直線または点線より長い横方向の第3の直線または点線が存在しない場合、前記第1の直線または点線及び前記第2の直線または点線を、横枠線の組として登録することにより、枠の外形が矩形でなく、且つ、点線が混在している場合にも枠の抽出が可能となる。

【0226】また、請求項32の発明によれば、2本の横方向の上側の第1の直線または点線及び下側の第2の直線または点線において、前記第1の直線または点線より前記第2の直線または点線の方が短く、且つ、前記第2の直線または点線の下方に前記第1の直線または点線と同じ長さ、又は前記第1の直線または点線より長い横方向の第3の直線または点線が存在する場合、前記第1の直線または点線及び前記第3の直線または点線を、横枠線の組として登録することにより、枠の外形が矩形でなく、且つ、点線が混在している場合にも枠の抽出が可能となる。

【0227】また、請求項33の発明によれば、2本の横方向の上側の第1の直線または点線及び下側の第2の直線または点線において、前記第1の直線または点線に対して前記第2の直線または点線が右方向又は左方向に短く、且つ、前記第2の直線または点線の下方に前記第1の直線または点線と同じ長さ又は前記第1の直線または点線より長い横方向の第3の直線または点線が存在しない場合、前記第1の直線または点線及び前記第2の直線または点線を、横枠線の組として登録する。このことにより、枠の外形が矩形でなく、且つ、上側の直線または点線と下側の直線または点線とが横方向にずれて配置されている場合にも枠の抽出が可能となる。

【0228】また、請求項34の発明によれば、2本の

横方向の上側の第1の直線または点線及び下側の第2の直線または点線において、前記第1の直線または点線に対して前記第2の直線または点線が右方向又は左方向に短く、且つ、前記第2の直線または点線の下方に前記第1の直線または点線と同じ長さ又は前記第1の直線または点線より長い横方向の第3の直線または点線が存在する場合、前記第1の直線または点線及び前記第3の直線または点線を、横枠線の組として登録することにより、枠の外形が矩形でなく、且つ、上側の直線または点線と下側の直線または点線とが横方向にずれて配置されている場合にも枠の抽出が可能となる。

【0229】また、請求項35の発明によれば、2本の横方向の上側の第1の直線または点線及び下側の第2の直線または点線において縦方向の重なりがない場合、前記第2の直線または点線に代えて前記第2の直線または点線の下方に存在する横方向の第3の直線または点線を、横枠線の候補とすることにより、枠の外形が矩形でなく、且つ、上側の直線または点線と下側の直線または点線とが縦方向にずれて配置されている場合にも枠の抽出が可能となる。

【0230】また、請求項36の発明によれば、下側の横枠が右方向へ長い又は左方向へ長い場合、横枠線の両方を横枠線の候補として、再度枠判断処理のために選択可能とすることにより、枠の外形が矩形でなく、且つ、上側の直線または点線と下側の直線または点線とが横方向にずれて配置されている場合においても、未処理の横枠の組を作ることが可能となる。

【0231】また、請求項37の発明によれば、長い方の横枠を横枠の候補として再度枠判断処理のために選択可能とすることにより、このことにより、上から順に横枠の組を作った場合において、未処理として残った横線または横点線を横枠の組とすることが可能となる。

【0232】また、請求項38の発明によれば、未処理の横枠の候補が存在している場合、再び部分パターンの上部から未処理の横枠を2本づつ組として枠の候補を抽出することにより、このことにより、上から順に横枠の組を作った場合において、未処理として残った横線または横点線を横枠の組とすることが可能となる。

【0233】また、請求項39の発明によれば、2本の横枠に両端が達している縦線を縦枠として登録することにより、直線と点線とが混在する縦枠を抽出する場合において、文字と枠とを正確に区別することができる。

【0234】また、請求項40の発明によれば、点線を抽出する際、構成する点線間の間隔に応じて、原画像の両端の点の位置を外延させて点線を表現する矩形の範囲を決定することにより、より正確に点線の位置を記述することが可能となる。

【0235】また、請求項41の発明によれば、点線が枠を構成する直線であるか否かを判断して矩形枠を抽出することにより、横点線の長さが異なっている場合にお

いても、横点線の組を作ることができ、枠の外形が矩形状となっていない点線で構成された表の枠を正確に抽出できる。

【0236】また、請求項42の発明によれば、横直線を上から順番に2本ずつ組にして横枠線とすることにより、横直線の長さが異なっている場合においても、横直線の組を作ることができ、枠の外形が矩形状となっていない表の枠を正確に抽出できる。

【0237】また、請求項43の発明によれば、連結した部分パターンから直線又は点線を抽出し、抽出した直線又は点線が枠あるかどうかを判断し、その枠の情報に基づいて文字領域を抽出することにより、様々な構造の表形式枠における文字領域を正確に抽出することが可能となる。

【0238】また、請求項44の発明によれば、横直線を上から順番に2本ずつ組にして横枠とすることにより、横直線の長さが異なっている場合においても、横直線の組を作ることができ、枠の外形が矩形状となっていない表の枠を正確に抽出できる。

【図面の簡単な説明】

【図1】本発明の一実施例によるパターン抽出装置の概略構成を示すブロック図である。

【図2】本発明の一実施例によるパターン抽出装置の連結パターン抽出処理を示す図である。

【図3】本発明の一実施例によるパターン抽出装置のマスク処理により横線分を抽出する処理を示す図である。

【図4】本発明の一実施例によるパターン抽出装置のマスク処理の動作を示すフローチャートである。

【図5】本発明の一実施例によるパターン抽出装置のマスク処理を示す図である。

【図6】本発明の一実施例によるパターン抽出装置の直線抽出処理における隣接投影法を説明する図である。

【図7】本発明の一実施例によるパターン抽出装置の直線抽出処理における隣接投影法の動作を示すフローチャートである。

【図8】本発明の一実施例によるパターン抽出装置の線分抽出処理を示す図である。

【図9】本発明の一実施例によるパターン抽出装置の線分検出処理の動作を示すフローチャートである。

【図10】本発明の一実施例によるパターン抽出装置の線分統合処理を説明する図である。

【図11】本発明の一実施例によるパターン抽出装置の線分統合処理の動作を示すフローチャートである。

【図12】本発明の一実施例によるパターン抽出装置の直線抽出処理を示す図である。

【図13】本発明の一実施例によるパターン抽出装置の直線探索処理を示す図である。

【図14】本発明の一実施例によるパターン抽出装置の横直線探索処理の動作を示すフローチャートである。

【図15】本発明の一実施例によるパターン抽出装置の

横直線探索処理における検索方向を示す図である。

【図16】本発明の一実施例によるパターン抽出装置の点線抽出処理を示す図である。

【図17】本発明の一実施例によるパターン抽出装置の点線抽出処理を示す図である。

【図18】本発明の一実施例によるパターン抽出装置の点線抽出処理を示す図である。

【図19】本発明の一実施例によるパターン抽出装置の点線抽出処理を示す図である。

10 【図20】本発明の一実施例によるパターン抽出装置の点線抽出処理を示す図である。

【図21】本発明の一実施例によるパターン抽出装置の点線抽出処理の動作を示すフローチャートである。

【図22】本発明の一実施例によるパターン抽出装置の直線及び点線抽出処理で抽出された直線及び点線を示す図である。

【図23】本発明の一実施例によるパターン抽出装置の矩形表現処理を示す図である。

20 【図24】本発明の一実施例によるパターン抽出装置の横直線の組の決定処理の動作を示すフローチャートである。

【図25】本発明の一実施例によるパターン抽出装置の横直線の組の決定処理の動作を示すフローチャートである。

【図26】本発明の一実施例によるパターン抽出装置の横直線の組の決定処理における横直線の配置状態を示す図である。

【図27】本発明の一実施例によるパターン抽出装置の横線の組の決定方法を示す図である。

30 【図28】本発明の一実施例によるパターン抽出装置の横線の組の決定方法を示す図である。

【図29】本発明の一実施例によるパターン抽出装置の横線の組の決定方法を示す図である。

【図30】本発明の一実施例によるパターン抽出装置の横線の組の決定方法を示す図である。

【図31】本発明の一実施例によるパターン抽出装置の入れ子構造の枠の決定方法を示す図である。

【図32】本発明の一実施例によるパターン抽出装置の縦枠決定処理及び入れ子処理の動作を示すフローチャートである。

【図33】本発明の一実施例によるパターン抽出装置の横枠決定処理を示す図である。

【図34】本発明の一実施例によるパターン抽出装置の縦枠決定処理を示す図である。

【図35】本発明の一実施例によるパターン抽出装置の矩形範囲抽出処理を示す図である。

【図36】本発明の一実施例によるパターン抽出装置の入れ子構造の処理を示す図である。

50 【図37】本発明の一実施例によるパターン抽出装置の点線抽出処理を示す図である。

【図38】本発明の一実施例によるパターン抽出装置の枠領域の決定方法を示す図である。

【図39】本発明の一実施例によるパターン抽出装置の対象外とするパターンを示す図である。

【図40】本発明の一実施例によるパターン抽出処理の原画像を示す図である。

【図41】本発明の一実施例によるパターン抽出処理の処理後の画像を示す図である。

【図42】従来の文字認識装置で抽出可能な枠を示す図である。

【図43】従来の文字認識装置における一文字枠の抽出方法を示す図である。

【図44】従来の文字認識装置における横一行ブロック枠の抽出方法を示す図である。

【図45】従来の文字認識装置におけるフリーフォーマット枠の抽出方法を示す図である。

【図46】従来の文字認識装置における表形式ブロック枠の抽出方法を示す図である。

【図47】従来の文字認識装置で抽出不可能な枠を示す図である。

【符号の説明】

1 連結パターン抽出部

2 マスク処理部

3 横直線抽出部

4 横隣接投影部

5 横線分検出部

6 横線分統合部

7 横直線検出部

8 横直線探索部

9 横点線検出部

10 縦直線抽出部

11 縦隣接投影部

12 縦線分検出部

13 縦線分統合部

14 縦直線検出部

15 縦直線探索部

16 縦点線検出部

17 横枠決定部

18 縦枠決定部

19 入れ子構造処理部

20 矩形内の横枠決定部

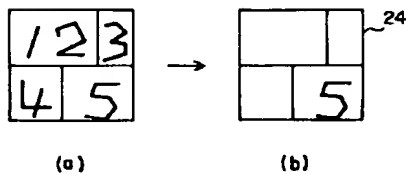
21 縦枠決定部

22 矩形表現部

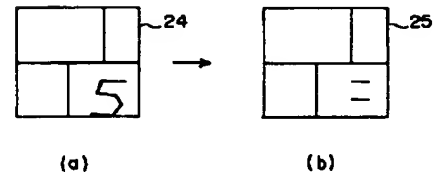
【図2】

【図5】

本発明の一実施例によるパターン抽出装置
の連結パターン抽出処理を示す図

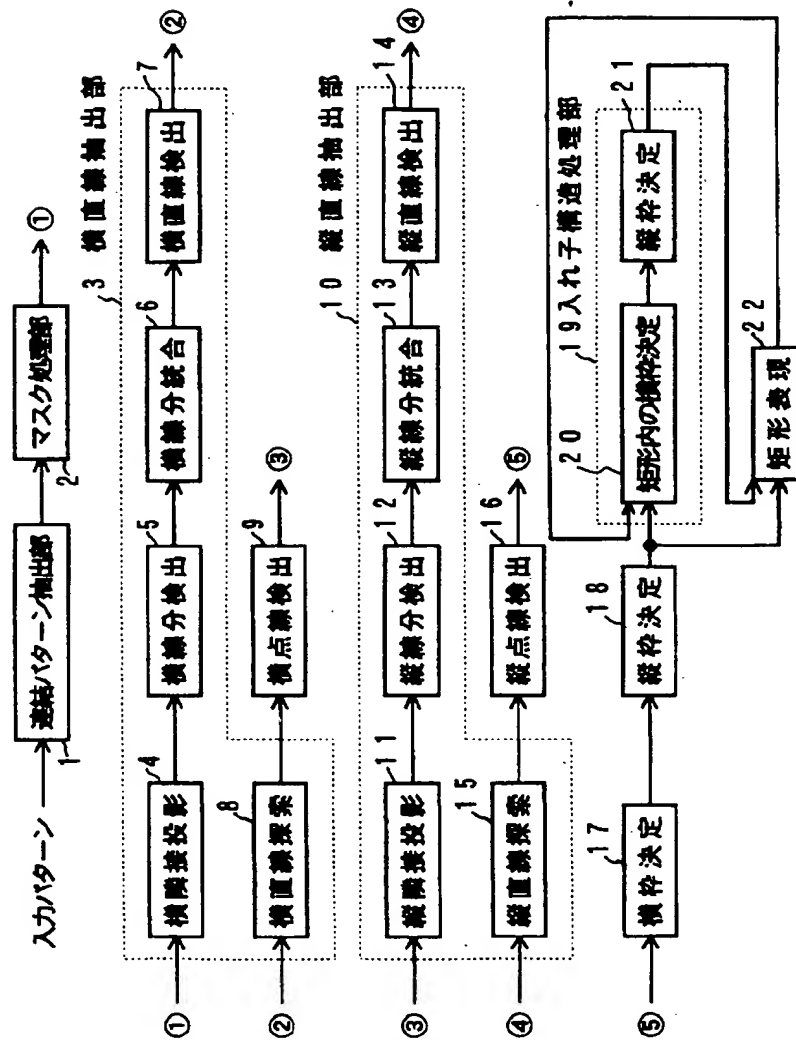


本発明の一実施例によるパターン抽出装置
のマスク処理を示す図



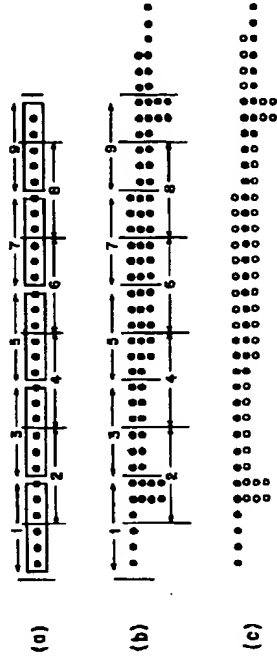
【図1】

本発明の一実施例によるパターン抽出装置の
概略構成を示すブロック図



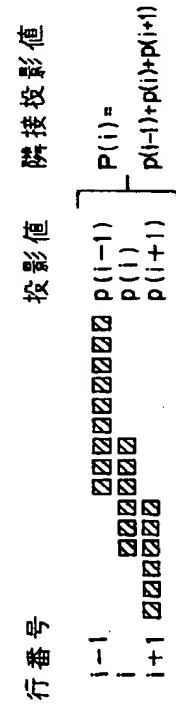
【図3】

本発明の一実施例によるパターン抽出装置
のマスク処理により横線を抽出する処理を示す図



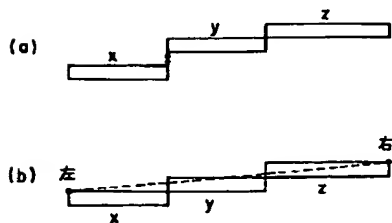
【図6】

本発明の一実施例によるパターン抽出装置の
直線抽出処理における隣接投影法を説明する図



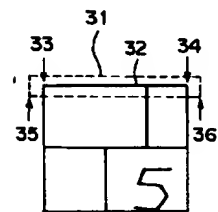
【図10】

本発明の一実施例によるパターン抽出装置の
線分統合処理を説明する図



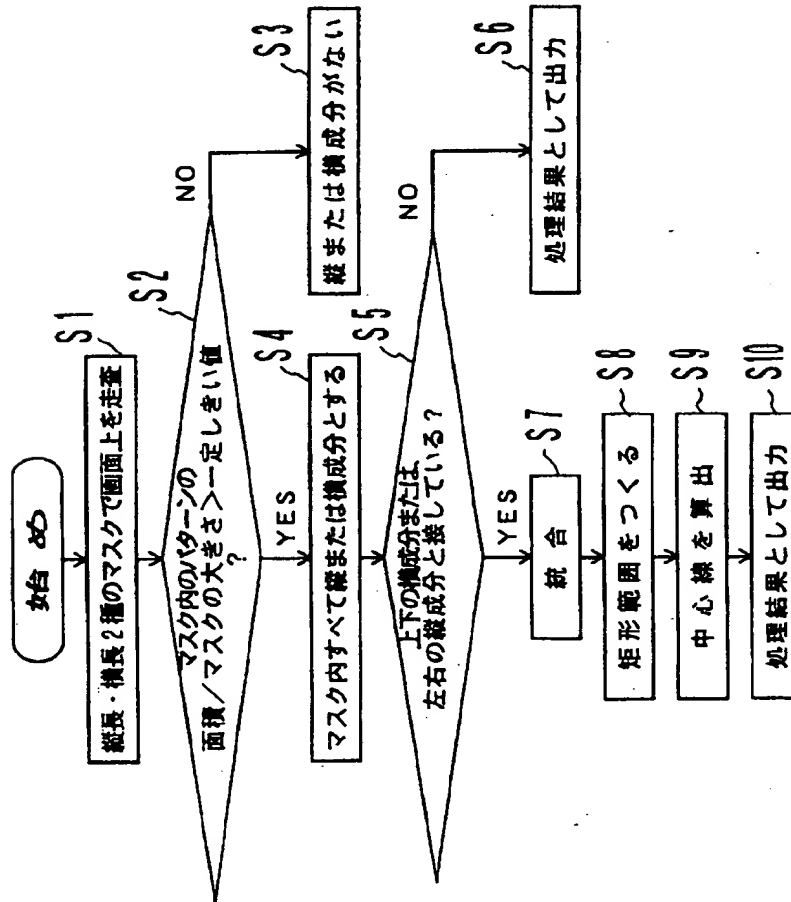
【図13】

本発明の一実施例によるパターン抽出装置の
直線探索処理を示す図



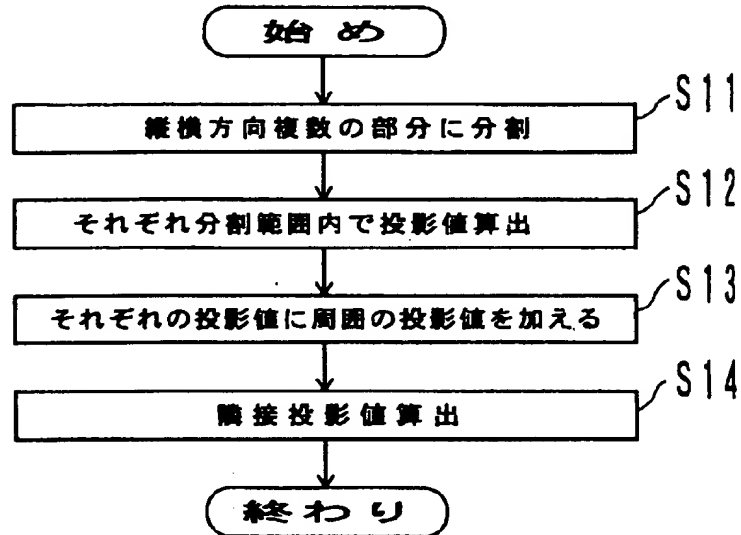
【図4】

本発明の一実施例による
パターン抽出装置のマスク処理の動作を
示すフローチャート



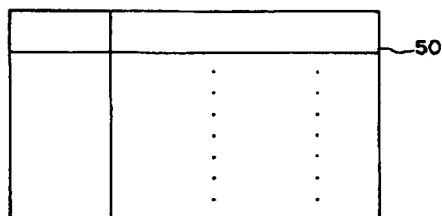
【図7】

本発明の一実施例によるパターン抽出装置の直線
抽出処理における隣接投影法の動作を示すフローチャート



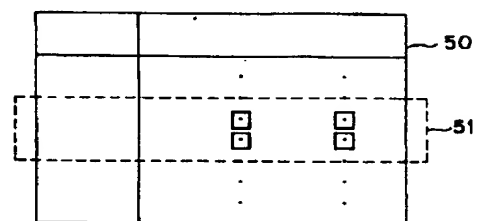
【図16】

本発明の一実施例によるパターン抽出装置の
点線抽出処理を示す図



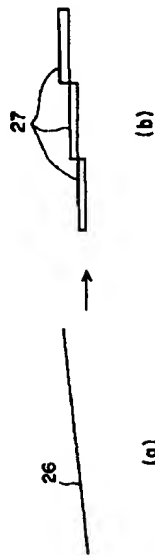
【図17】

本発明の一実施例によるパターン抽出装置の
点線抽出処理を示す図



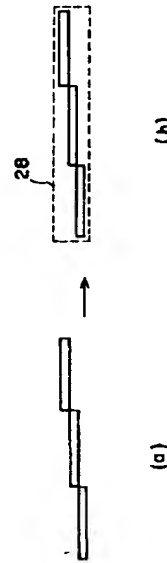
【図8】

本発明の一実施例によるパターン抽出装置の
線分抽出処理を示す図



【図12】

本発明の一実施例によるパターン抽出装置の
直線抽出処理を示す図



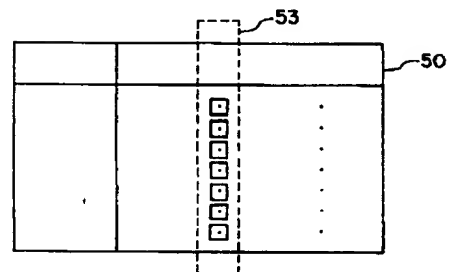
【図15】

本発明の一実施例によるパターン抽出装置の
横直線探索処理における探索方向を示す図

探索方向 位置	右方向	左方向
隣		
斜め隣		

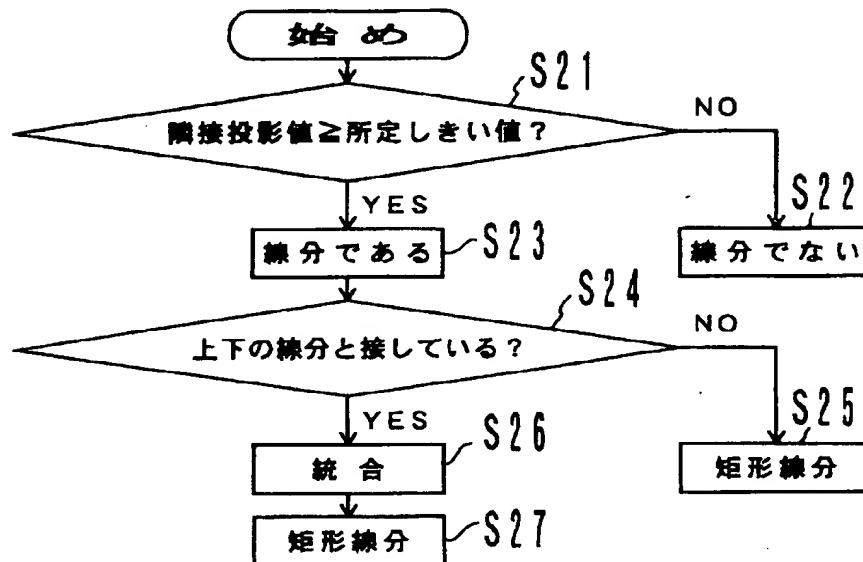
【図18】

本発明の一実施例によるパターン抽出装置の
点線抽出処理を示す図



【図9】

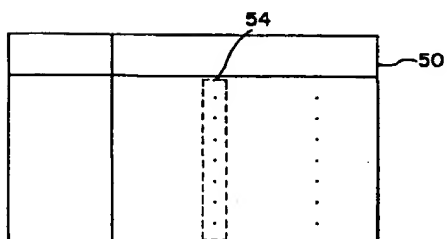
本発明の一実施例によるパターン抽出装置の
線分検出処理の動作を示すフローチャート



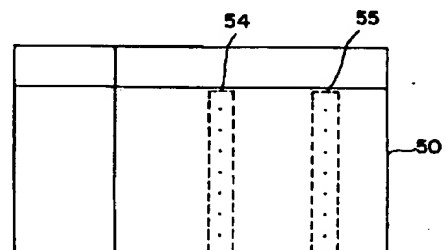
【図19】

【図20】

本発明の一実施例によるパターン抽出装置の
点線抽出処理を示す図

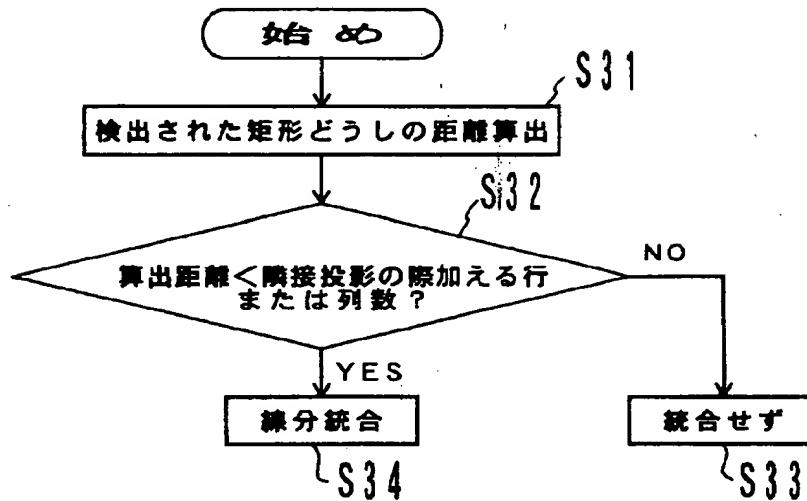


本発明の一実施例によるパターン抽出装置の
点線抽出処理を示す図



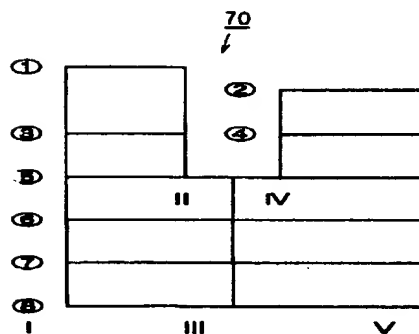
【図11】

本発明の一実施例によるパターン抽出装置の
線分統合処理の動作を示すフローチャート



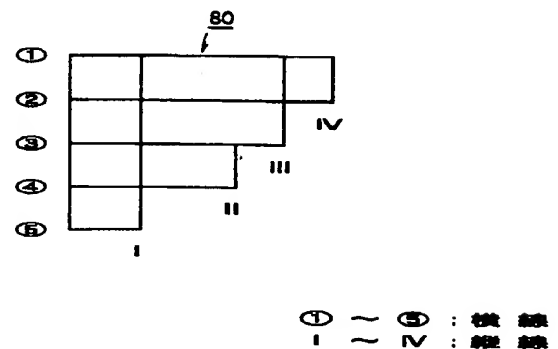
【図27】

本発明の一実施例によるパターン
抽出装置の横線の組の決定方法を示す図



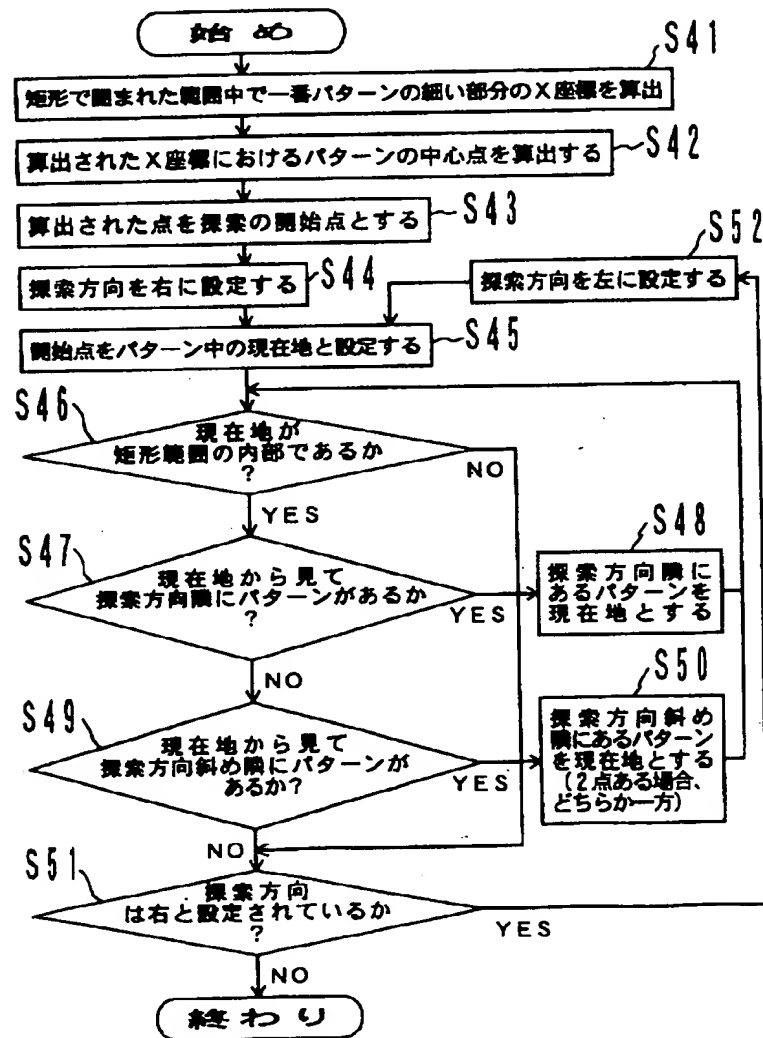
【図28】

本発明の一実施例によるパターン
抽出装置の横線の組の決定方法を示す図



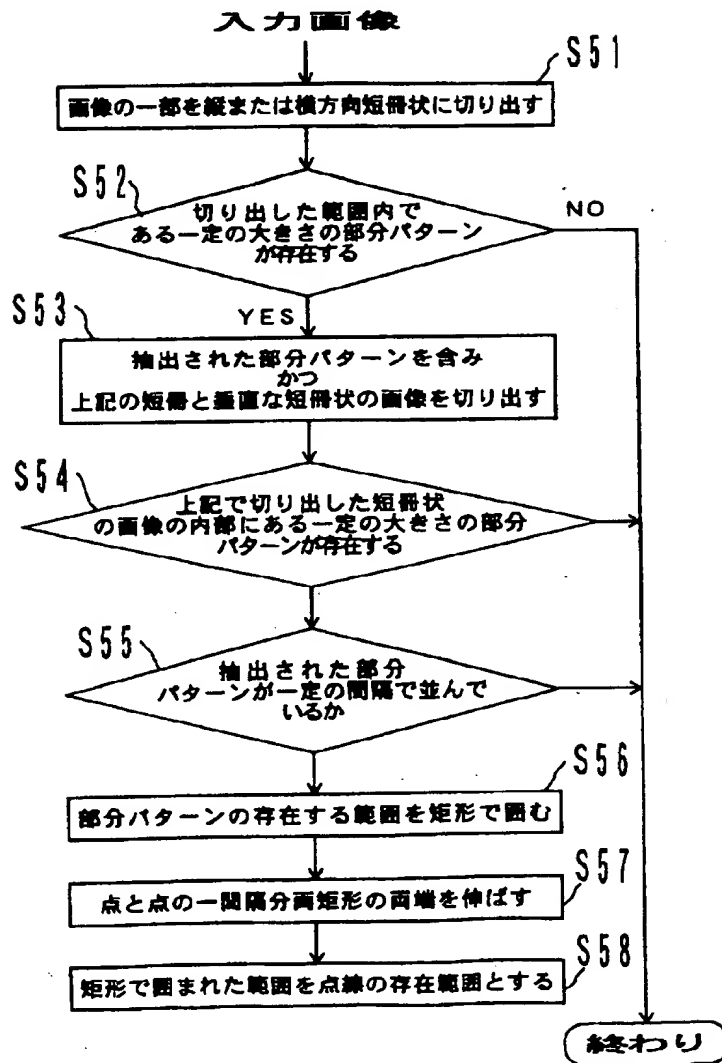
【図14】

本発明の一実施例によるパターン抽出装置の
横直線探索処理の動作を示すフローチャート



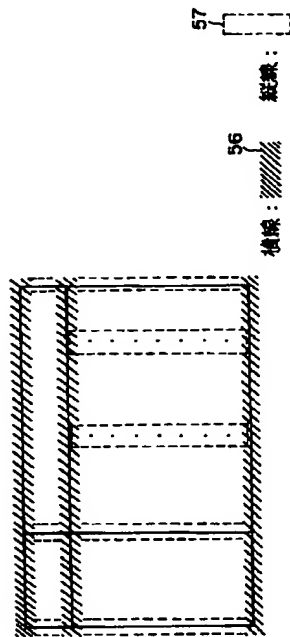
【図21】

本発明の一実施例によるパターン抽出装置の
点線抽出処理の動作を示すフローチャート



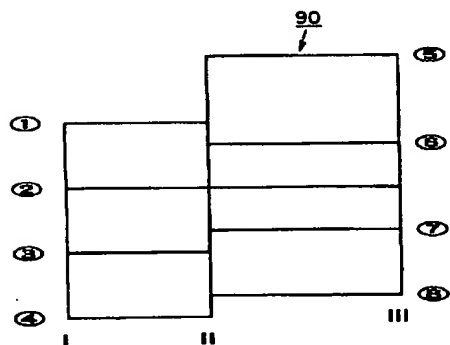
【図22】

本発明の一実施例によるパターン抽出装置の直線及び点線抽出処理で抽出された直線及び点線を示す図



【図29】

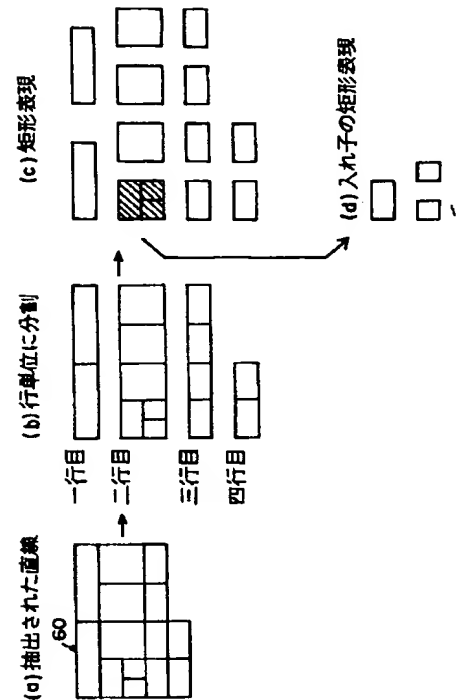
本発明の一実施例によるパターン抽出装置の横線の組の決定方法を示す図



① ~ ⑧ : 横線
I ~ III : 縦線

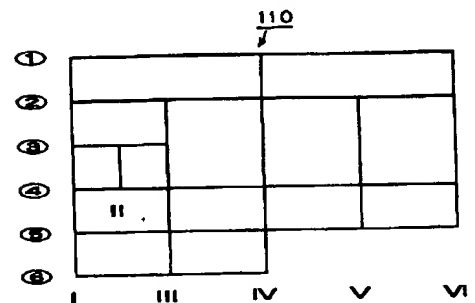
【図23】

本発明の一実施例によるパターン抽出装置の矩形表現処理を示す図



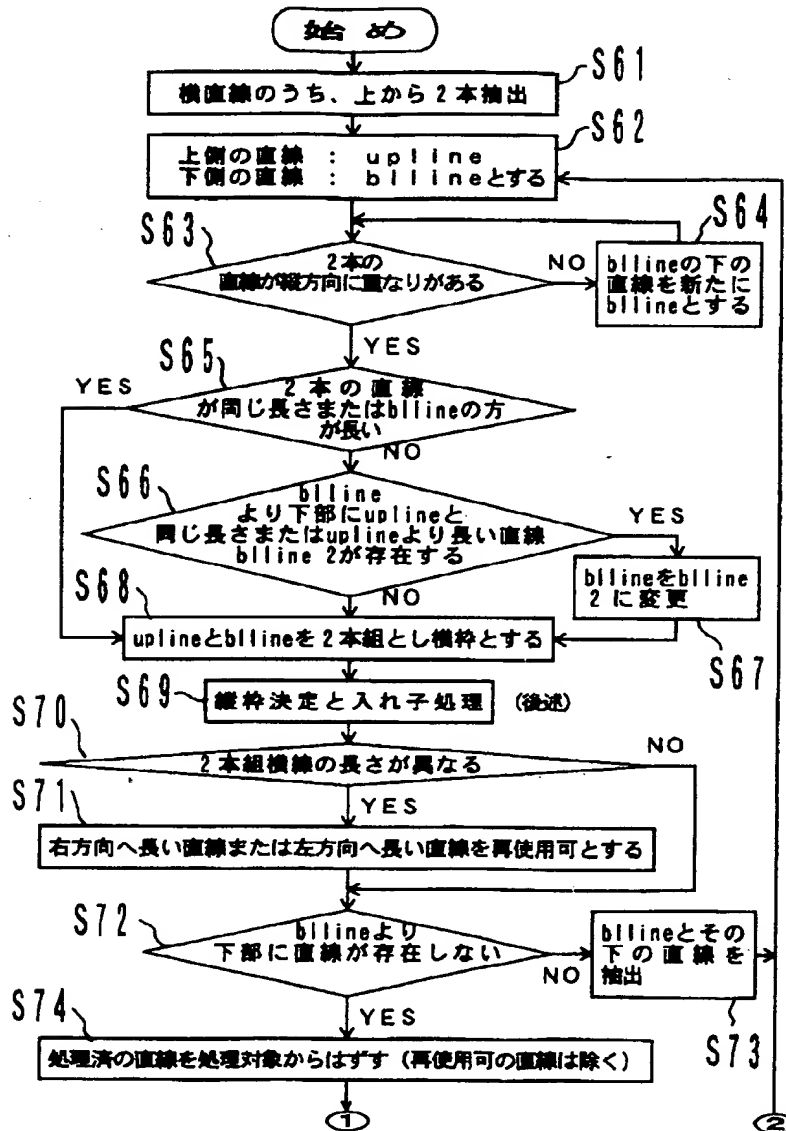
【図33】

本発明の一実施例によるパターン抽出装置の横線決定処理を示す図



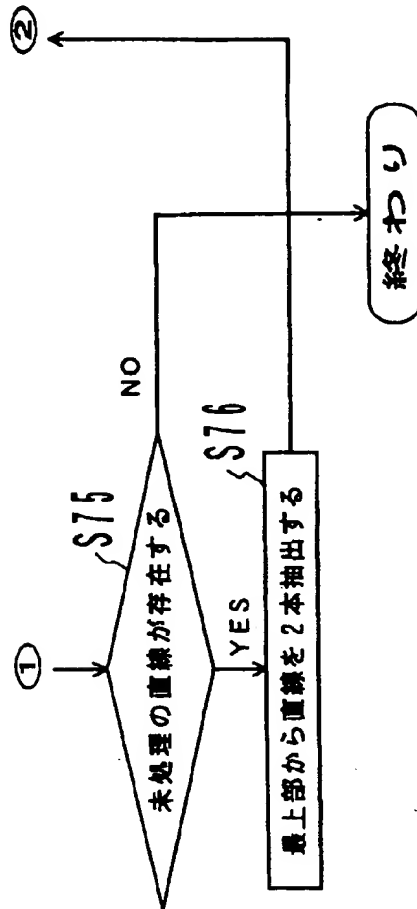
【図24】

本発明の一実施例によるパターン抽出装置の
横直線の組の決定処理の動作を示すフローチャート



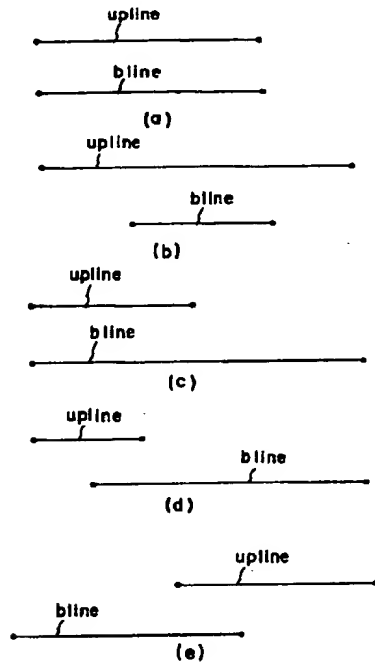
【図25】

本発明の一実施例によるパターン抽出装置の
横直線の粗の決定処理の動作を示すフローチャート



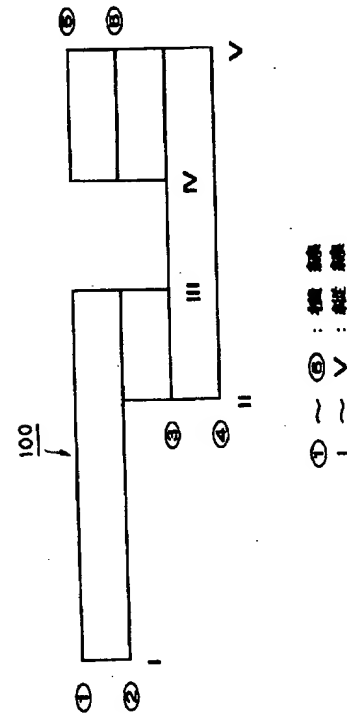
【図26】

本発明の一実施例によるパターン抽出装置の横直線の組の決定処理における横直線の配置状態を示す図



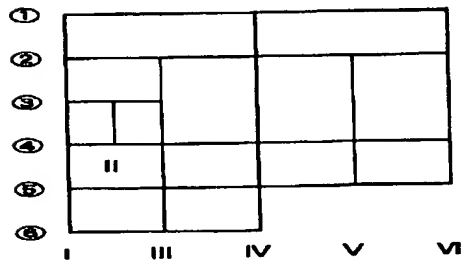
【図30】

本発明の一実施例によるパターン抽出装置の横線の組の決定方法を示す図



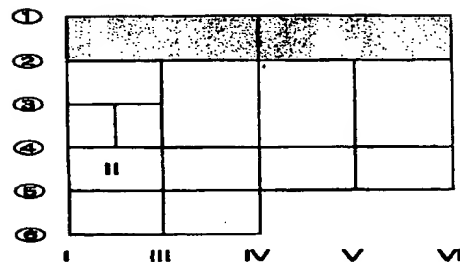
【図34】

本発明の一実施例によるパターン抽出装置の縦枠決定処理を示す図



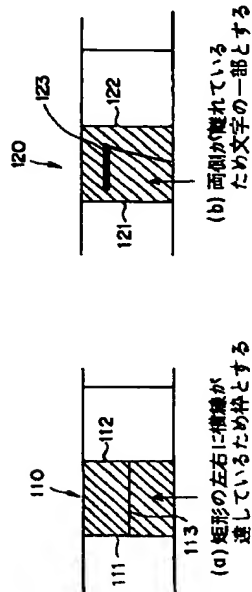
【図35】

本発明の一実施例によるパターン抽出装置の矩形範囲抽出処理を示す図



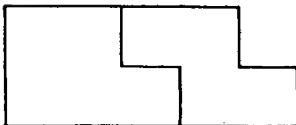
【図31】

本発明の一実施例によるパターン抽出装置の
入れ子構造の枠の決定方法を示す図



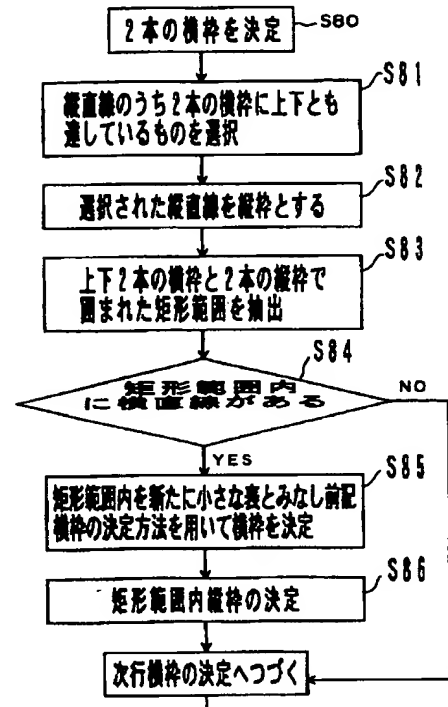
【図39】

本発明の一実施例によるパターン抽出装置
の対象外とするパターンを示す図



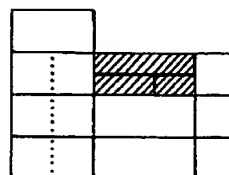
【図32】

本発明の一実施例によるパターン抽出装置の縦枠決定処理
及び入れ子処理の動作を示すフローチャート



【図47】

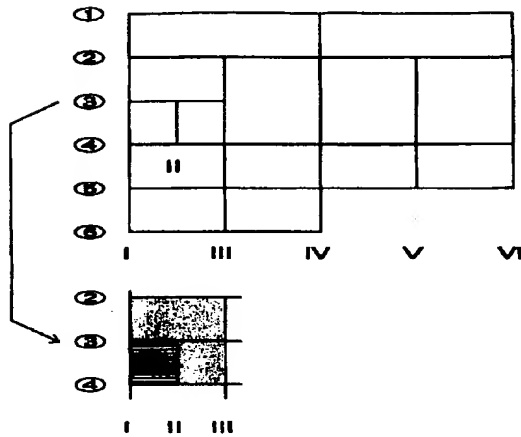
従来の文字認識装置で抽出不可能な枠を示す図



【図36】

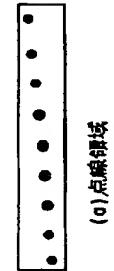
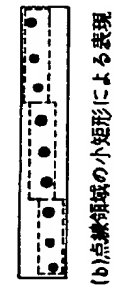
【図37】

本発明の一実施例によるパターン
抽出装置の入れ子構造の処理を示す図

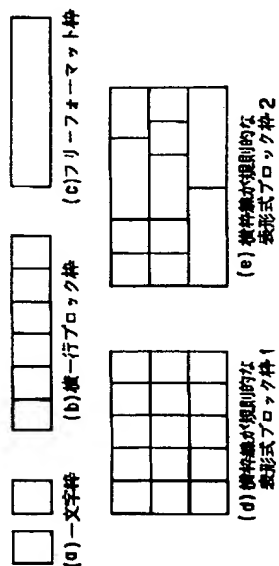


【図4 2】

本発明の一実施例によるパターン抽出装置の
点線抽出処理を示す図



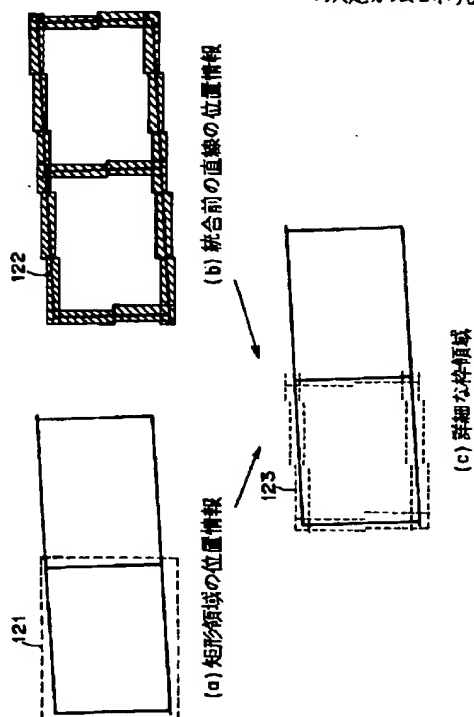
従来の文字認識装置で抽出可能な枠を示す図



【图 3 8】

【図40】

本発明の一実施例によるパターン抽出装置の枠領域
の決定方法を示す図



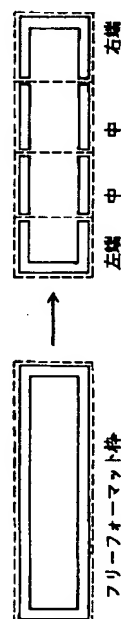
本発明の一実施例によるパターン抽出処理の原画像を示す図

<div style="display: flex; justify-content: space-between; align-items: center;"> 140 12345 </div>			
~清得係分組について~		会場名: FUJITSU LAB 000 記入者: ATUKO 会期日: 平成 7年 3月28日	
要 任 者		開催期 日	1: 机上 2: 室内 3: 廊下 4: ロビー 5: 他 6: その他
担当	氏 名	担当	氏 名
1) ほうこう	高士 通夫	2) 高橋	高士 通夫
3) はたき	高士 通夫	3) 高橋	高士 通夫
4) 藤村	高士 通夫	4) 高橋	高士 通夫
5) 藤村	高士 通夫	5) 高橋	高士 通夫
6) 藤村	高士 通夫	6) 高橋	高士 通夫
7) 藤村	高士 通夫	7) 高橋	高士 通夫
8) 藤村	高士 通夫	8) 高橋	高士 通夫
9) 藤村	高士 通夫	9) 高橋	高士 通夫
会 場 内 容			
1: 立席 2: 座席 3: 廊下なし 4: ロビー 5: 他 6: その他 7: 室内あり 8: 室内なし 9: 廊下あり 10: 廊下なし			

開催: 当グループの清得係分組について、3月20日午後2時から開催を行った。決定事項は以下の通りである。ほうこう責任者高士通夫、高橋責任者高士通夫、はたき責任者高士通夫、こみね責任者高士通夫、藤村責任者高士通夫。当グループの清得係分組について、3月20日午後2時から開催を行った。決定事項は以下の通りである。ほうこう責任者高士通夫、高橋責任者高士通夫、はたき責任者高士通夫、こみね責任者高士通夫、藤村責任者高士通夫。当グループの清得係分組について、3月20日午後2時から開催を行った。決定事項は以下の通りである。ほうこう責任者高士通夫、高橋責任者高士通夫、はたき責任者高士通夫、こみね責任者高士通夫、藤村責任者高士通夫。

【图45】

従来の文字認識装置におけるフリーフォーマット枠の
抽出方法を示す図

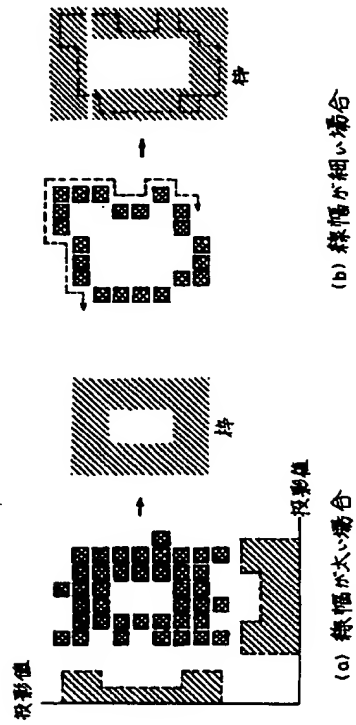


本発明の一実施例によるパターン抽出処理の処理後の画像を示す図

<div style="border: 1px solid black; padding: 2px; display: inline-block;">12345</div>				
<div style="border: 1px solid black; padding: 2px;"> 番 名 </div>	<div style="border: 1px solid black; padding: 5px;"> ～清掃係分組について～ </div>	会 場 ： FUJITSU LAB 000 日 ： ATUKO 会 期 ： 平成 7年 3月20日		
<div style="border: 1px solid black; padding: 5px; display: inline-block;">責 任 者</div>		<div style="display: flex; justify-content: space-between;"> <div style="border: 1px solid black; padding: 2px;"> 機上 期 日 </div> <div style="border: 1px solid black; padding: 2px;"> 1: 机 2: 上 3: 下 決り 決り </div> <div style="border: 1px solid black; padding: 2px;"> 2: 室内 4: ロビー 5: 廊下 6: 門内有効 </div> </div>		
<div style="border: 1px solid black; padding: 2px;">組 当</div>	<div style="border: 1px solid black; padding: 2px;">氏 名</div>	<div style="border: 1px solid black; padding: 2px;">組 当</div>	<div style="border: 1px solid black; padding: 2px;">氏 名</div>	<div style="border: 1px solid black; padding: 5px; text-align: center;">会 議 内 容</div>
1) ほうき 2) ほうき 3) ほうき 4) ほうき 5) ほうき 6) ほうき 7) ほうき 8) ほうき 9) ほうき	1) ほうき 2) ほうき 3) ほうき 4) ほうき 5) ほうき 6) ほうき 7) ほうき 8) ほうき 9) ほうき	2) ほうき 3) ほうき 4) ほうき 5) ほうき 6) ほうき 7) ほうき 8) ほうき 9) ほうき	1) ほうき 2) ほうき 3) ほうき 4) ほうき 5) ほうき 6) ほうき 7) ほうき 8) ほうき 9) ほうき	<div style="border: 1px solid black; padding: 5px;"> 1: 資料あり 2: 資料なし 3: 資料あり 4: 資料なし 5: 資料あり 6: 資料なし 7: 資料あり 8: 資料なし 9: 資料あり </div>
要 要 : 当グループの清掃係分組について、3月20日午後2時から会議を行った。決り事項は以下の通りである。ほうき責任者富士通夫、ほうき責任者富士通子、ほうき責任者富士通太、ほうき責任者富士通美、ほうき責任者富士通雄。当グループの清掃係分組について、3月20日午後2時から会議を行った。決り事項は以下の通りである。ほうき責任者富士通夫、ほうき責任者富士通子、ほうき責任者富士通太、ほうき責任者富士通美、ほうき責任者富士通雄。当グループの清掃係分組について、3月20日午後2時から会議を行った。決り事項は以下の通りである。ほうき責任者富士通夫、ほうき責任者富士通子、ほうき責任者富士通太、ほうき責任者富士通美、ほうき責任者富士通雄。				
機 上				
室内1				
室内2				
廊 下 1				
廊 下 2				
ロビー				
1) 8:00 2) 14:00 3)	1) 9:00 2) 15:00 3)	1) 10:00 2) 16:00 3)	1) 11:00 2) 17:00 3)	1) 12:00 2) 18:00 3)
1報告必要 ②不必要				

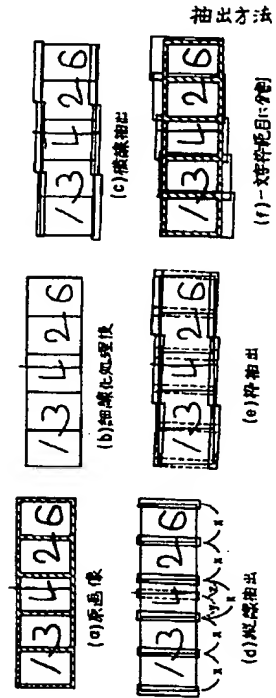
【図43】

従来の文字認識装置における
一文字枠の抽出方法を示す図



【図44】

従来の文字認識装置における横一行ブロックの
抽出方法を示す図



【図46】

従来の文字認識装置における表形式ブロック枠の
抽出方法を示す図

